International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 28575-28585

http://www.ripublication.com

Human Activity Analysis For Video Surveillance Based on Visual Features

[1]Sivaprakash Paramasivan and [2]Ravichandran C Gopalakrishnan
[1]Assistant Professor, Department of ECE, RVS College of Engg. and Tech., Dindigul, Tamilnadu, India.
Sivabtl85@gmail.com
[2]Principal, SCAD Institute of Technology, Palladam, Tamilnadu, India.
cgravichandran@gmail.com

Abstract

Motion can defined as to detect through measure change in speed or vector of an object or objects. Motion detection is extensively used in various ways in video surveillance application and to detect and track the human and human activities in real time video sequence. This paper is related to the extensive subject of motion detection and analysis in video surveillance of image sequence. Now a day in video surveillance application is used to detect multiple objects and monitor their activities are challenging task in indoor and outdoor environment. Consider spatio-temporal relationship among feature points, thereby enabling detection and classification of simple and complex human activities. In presence of a good number of real time problems such as, the problems are namely illumination changes, moving background and shadow detection. The robust algorithm is proposed for enhancing the accuracy and reliability of motion detection and classification methods to develop the real time video sequence in video surveillance. Its advantages are discussed and compared to the relative approaches for action recognition. The widely-used KTH human activity dataset demonstrated and they are the implemented state-of-the art methods.

Keywords: motion analysis, Spatio-temporal features, video surveillance, action recognition, and video monitoring.

Introduction

Motion analysis is an important task within the field of computer vision. Human motion analysis addresses general common tasks, such as: person detection and tracking, activity classification, behavior interpretation and also person identification.

Recognition is the identification of objects in an image. This process would probably start with image processing techniques such as noise removal, followed by (low-level) feature extraction to locate lines, regions and alignment areas with certain textures. 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. Human activity recognition has become an important area of research in computer vision in recent years. Recently, local spatio-temporal features are extensively used in action recognition tasks, because they are more robust to noise, occlusion, and geometric variations than global (or large-scale) features. It has gained a lot of attention because of its important application domains like video indexing, surveillance, human computer interaction, sport video analysis, intelligent environments etc. They have been developed into different types of algorithms for human activity recognition [4]-[20]. They discussed recognition algorithms in detail and apply them to specific groups of activities, including single person interaction, multiple interactions, and person vehicle interactions. Brief over view of core technologies has been discussed [1]; techniques are often sensitive to poor resolution, frame rate, drastic illumination change in surveillance systems. Probability based methods for activity recognition from trajectories has been addressed using statistical models, e.g., hidden Markov models (HMMs),continuous random filed(CRFs),skip-chain(SCCRFs) [2],[3] published a topical review of activity recognition research, this extensive review focuses on activity recognition techniques in which only on-body sensors are used and is also aimed at score functions of activity recognition. The EPs (Emerging patterns) based classifier model uses a set of multi-attribute tests for each class of activity [3]. Consider the blob features including mean, variance of each blob, luminance. They combined the non-zero pixels in the feature image into blobs using connected component analysis method. Based on the blob appearance the activities can be recognized [4].They proposed human detector framework based on HOG features in state of art field detection systems[16] and they defined a new way of computing HOG features based on square block of histograms which is four time faster than[18] implementation based on the integral of histograms. They exploited this increased speed to apply an approximate Gaussian mask to the features in order to improve the feature quality and the corresponding detection rate. Histogram of gradients features are used to detect whether an image encompasses human beings or not. SVM was used to train the classifier on the features [17].They are considered as the triaxial accelerometer signals, the accelerometer is used to collect human motion acceleration data for classifying the different type of human activities[8][12]. The low calculation cost feature extension method for 3D accelerometer signals is used in human activity recognition. Haar like feature is used to state of the art face detector, for high performance and calculation efficiency [8]. They introduced FIS (Fuzzy Inference system) system andit can distinguish the motion patterns of its ability of decision making. Three different features including peak to peak amplitude, standard deviation, and correlation between axes are extracted from each axis of the accelerator as inputs to the fuzzy system [12]. They proposed fuzzy rule based approach to recognize human activities. Consider the motion-based and shape based features to recognize an activity, many activities remain unidentified as temporal information is discarded. They motivated to design a robust method that uses temporal information [13]. The recognition algorithm is based on the characterizing behavior in terms of spatiotemporal features [19], [6], [20], [14], and [7]. They have been implemented a new spatiotemporal detector and a number of cuboid descriptors were analyzed. The cuboid is extracted which contains the spatio-temporally windowed pixel values and then capable of dealing with low resolution and noisy data. Cuboid prototypes by clustering a large number of cuboids are extracted from the training data set, cluster using k-means algorithm. A cluster prototype is a very simple yet powerful method for reducing variability of the data while maintaining its richness [19].They proposed frame differencing method only for simple low-level visual features and motion captured ,recognition can be achieved accurately using optical flow method. Here introduced a novel weighted-sequence distance (WSD) measure for comparing the similarity between two sequences. Support vector machine (SVM) is used for classifying the activity in a coded video sequence. This approach considers both global and local spatial temporal structures. It does not include multiple actions/subjects and appearance variations due to view angle changes [6]. They are considered as local space-time features that capture local events in video sequence, frequency and moving patterns. Recognizing human action directly from image measurements, the image measurements in terms of optic flow or spatio-temporal gradients, recognition can be achieved using local measurements in terms of spatiotemporal interest points; Support Vector Machines (SVMs) are state-of-the-art large margin classifiers that combined with motion descriptors in terms of local features (LF) and feature histograms (HistLF)[20].They implemented the probability based models to handle noisy feature points arisen from dynamic background and moving camera. This is achieved by using latent topic models such as the probabilistic Latent Semantic Analysis (PLSA) model and Latent Dirichlet Allocation (LDA). The proposed algorithm can also localize multiple actions in complex motion sequences containing multiple actions. Both PLSA and LDA methods give higher recognition performance [14].Multi-stage approach is proposed to detect and recognize the human activities. This approach is to identify and extract salient space-time volumes that exhibit smooth periodic motion. They presented an algorithm that accelerates sparse coding by recursively constructing basis vectors, adjusting only a fraction of the weights at any given time. They are implemented to more challenging video collections that feature a small range of background, significant clutter and intermittent occlusion only [7].They modeled sub event (primitive) detectors of a spatiotemporal model for sequential event changes. Specialized Viterbi algorithm is designed to learn and interference the targets sequential events and handle the event overlap simultaneously, for repetitive sequential human activities [5].

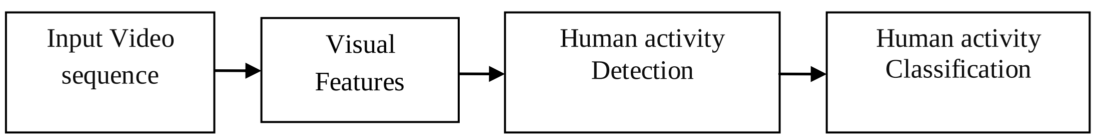

Figure 1: Block diagram of the Video Surveillance System

They introduced the systematic model based approach to learn the nature of such temporal variations (time warps). This approach allows us to learn the space of time warps for each activity while simultaneously capturing other intra and inter class variations [11].

This paper is organized as follows: Section1 provides the introduction.Section2 explains visual features for action detection. Section3 deals with the comparative analysis of existing classifiers. Section4 concludes the paper.

Visual Features For Action Detection

Now a day, the growth of video-based human action detection technology has been reached its heights. The extraction of appropriate features is critical for action detection. Ideally, visual features are able to face the following challenges for robust action detection:

1. Viewpoint variations of the camera.
2. Performing speed variations for different people.
3. Different anthropometry of the performers and their movement style variations.
4. Cluttered and moving backgrounds.

In the earlier days, human actions were tracked and segmented from the videos to characterize actions and motion trajectories are popularly used to represent and recognize actions. Unfortunately, only limited success has been achieved because robust object tracking itself is a nontrivial task. Recently, interest point based video features show promising results in the action detection research. Such interest point-based video features do not require any foreground/background separation or human tracking. The resources for four types of interest-point based features are listed below:

(A). Space-Time Interest Point (Stip)

The features of space-time interest point (STIP) have been frequently used for action recognition. However, the detected interest points are usually quite sparse, and it is time consuming to extract STIP features for high-resolution videos.

(B). Scale-Invariant Spatiotemporal Interest Point (SISTIP)

The next type of interest point features is called dense and scale-invariant spatiotemporal interest point (SISTIP), is compared to that of STIP features, the SSI-STIP features are scale-invariant (both spatially and temporally) and densely cover the video. The feature extraction is accelerated through the use of approximate box-filter operations on an integral video structure.

(C).Sparse Spatiotemporal Feature

The sparse spatiotemporal features are usually denser than the STIP features. However, they do not retain the features at multiple scales. When compared to the local feature detection, the proposed feature descriptor is relatively simple.

(D).Scale Invariant Feature Transformation (SIFT)
The 3-D SIFT descriptor is similar to the scale invariant feature transformation (SIFT) descriptor. But the gradient direction for each pixel is a three-dimensional vector. It can work with any interest point detector.

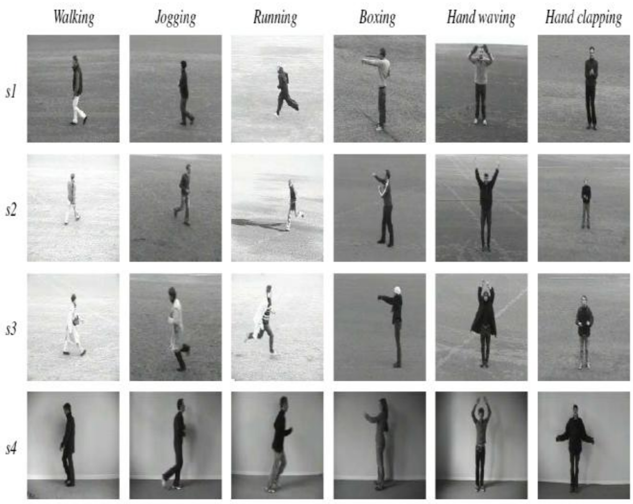

Figure 2: Example of KTH Dataset Consists of Six Activities

Comparative Analysis
In this section the brief review with some of the existing classification approaches has been given. Here, we discuss about existing classification approaches such as: NN Classifier, K-nn Classifier, Rule Based Classifier**,** Naive Bayes Classifier, SVM classifier as following sub divisions.

(A). NN Classifier:
In [6], [19] the nearest neighbor (NN) classifier is commonly used due to its simplicity and effectiveness. In 1-nn rule, an input is assigned to the class of its nearest neighbor from a stored labeled reference set. The goal of designing a NN classifier is to maximize the classification accuracy while minimizing the sizes of both the reference and feature sets. It has been recognized that the editing of the reference set and feature selection must be simultaneously determined when designing the NN classifier with high classification power. Let us consider a set of n patterns of known classification {x1, x2. . . x_n}, where it is assumed that each pattern belongs to one of the classes C_1, C_2, . . , C_i,. , C_K.

The NN classification rule that assigns a pattern x of unknown classification to the class of its nearest neighbor, where xi ∈ {x_1, x_2,… x_n} is defined to be the nearest neighbor of ‘x’ if

$$D\,(x_i; x) = \min \{D\,(x_l; x)\};\ l = 1, 2,.., n. \qquad (1)$$

Here D is any distance measure definable over the pattern space. Since the aforementioned scheme employs the class label of only the nearest neighbor to x, this is known as the 1-NN rule. On-parametric classifiers have several very important advantages such as: i) can naturally handle a huge number of classes. (ii) Avoid over fitting of parameters, which is a central issue in learning based approaches. (iii) Require no learning/training phase. The nearest neighbor rule is quite simple, but very computationally intensive and covers only small region in data set.

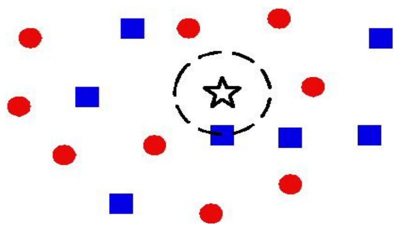

Figure 3: 1-Nn Classifier Model

(i). K-nn Classifier:

The k-nn classifier is a very simple non-parametric method for classification. Despite the simplicity of the algorithm, it performs very well and is an important standard method for classification. The k-nn classifier compute the distance between two classes using some distance function d(x, y), where x,y are classes composed N of features, such that x= $\{x_1, \ldots, x_N\}$, y=$\{y_1, \ldots, y_N\}$.

$$d_A(x, y) = \sum_{i=1}^{N} |x_i - y_i| \quad (2)$$

The traditional K-nn classification limitations as follows: great calculation complexity, fully dependent on training set, and no weight difference between each class.

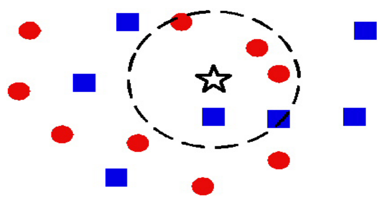

Figure 4: K-Nn Classifier Model

(B). Rule Based Classifier:

The Rule based classifier is one of the simplest classification methods. This approach utilizes a set of pre-determined rules to generate the final result. These rules consist of a collection of “if…then….” clauses.

Rule: (Condition) $\rightarrow$y

– where

- Condition is a conjunctions of attributes
- y is the class label

(C). Naive Bayes Classifier:

The Naive Bayes Classifier applies the Bayes theorem as the fundamental statistical principle when carrying out the classifications. Bayes theorem utilizes prior knowledge of the classes with new evidence gathered from subsequent input data. The following equation represents the compact form of the Bayes theorem which uses conditional probabilities:

$$p(C|F_1, \dots, F_n) \tag{3}$$

Over a dependent class variable *C* with a small number of outcomes or *classes*, conditional on several feature variables F_1 through F_n. The problem is that if the number of features *n* is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable. Using Bayes' theorem, we write

$$p(C|F_1,\dots,F_n) = \frac{p(C)\ p(F_1,\dots,F_n|C)}{p(F_1,\dots,F_n)}. \tag{4}$$

The discussion so far has derived the independent feature model, that is, the naive Bayes probability model. The naive Bayes classifier combines this model with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the maximum a posteriori or MAP decision rule. The corresponding classifier is the function classify defined as follows:

$$\mathrm{classify}(f_1,\dots,f_n) = \underset{c}{\mathrm{argmax}}\ p(C=c) \prod_{i=1}^{n} p(F_i = f_i|C=c). \tag{5}$$

An advantage of the Naive Bayes Classifier easy to implement, good results obtained in most of the cases. Disadvantages, Assumption: class conditional independence, therefore loss of accuracy and practically, dependencies exist among variables.

(D). Svm Classifier:

In [14], [20] the Support Vector Machine (SVM) is a classification technique, which was applied with great success in many challenging non-linear classification problems and was successfully applied to large data sets. SVM classification tasks due to their strong theoretical foundation and good classification accuracies that have been demonstrated in a wide range of application domains.

The SVM algorithm finds a hyperplane that optimally splits the training set. The optimal hyperplane can be distinguished by the maximum margin of separation between all training points and the hyperplane. Looking at a two-dimensional problem we actually want to find a line that “best” separates points in the positive class from points in the negative class. Generally w will be scaled by ||w||. The training part the algorithm needs to find the normal vector "w" that leads to the largest "*b*" of the hyperplane.For calculating the SVM, we see that the goal is to correctly classify all the data. For mathematical calculations we have,

[a] If Y_i= +1; $wx_i + b \geq 1$

[b] If Y_i= -1; $wx_i + b \leq 1$

[c] For all i; $y_i (w_i + b) \geq 1$

To understand the essence of SVM classification following four basic concepts such as:

(i) the separating hyperplane,
(ii) the maximum-margin hyperplane,
(iii) the soft margin and
(iv) the kernel function.

The most observable drawback to the SVM algorithm, as described thus far, is that it apparently only handles binary classification problems.

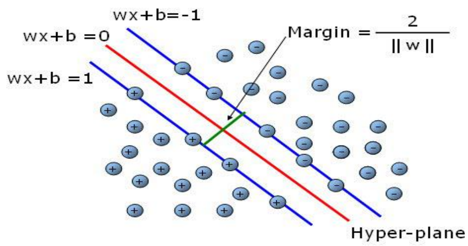

Figure 5: SVM Classifier Model

Table 1 compares the reported performance of several approaches of activity recognition on the KTH dataset.

Table 1: Represents the recognition accuracy to related work

Action	Box	Clap	Wave	Jog	Run	Walk	Mean
Niebles et al, 2008	98	86	93	53	88	82	83.3
Dean et al, 2009	81	80	86	69	89	81	81.1
Wang and Li, 2009	88	81	83	70	73	91	81
Doll´ar et al, 2005	80	82	84	63	73	89	78.5
Schuldt et al, 2004	98	60	74	60	55	84	71.8

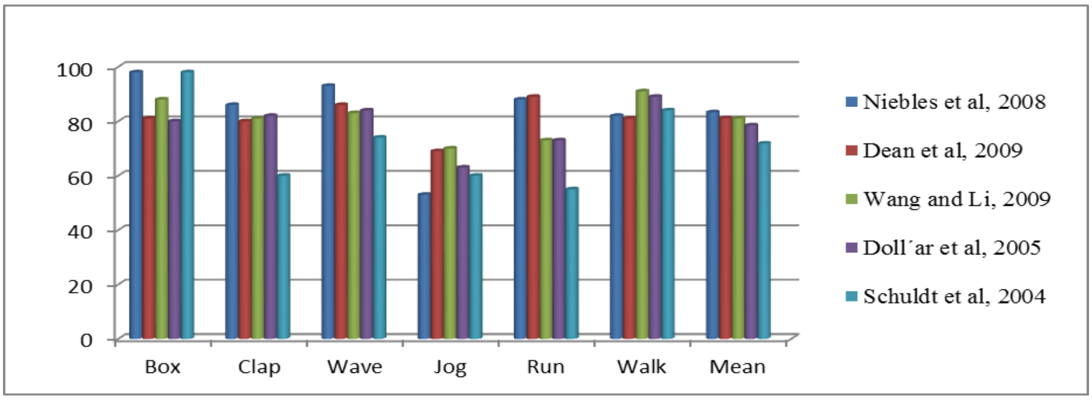

Figure 6: Represents The Several Approaches For The Human Activity Recognition Methods

Conclusion

This paper, gives the detailed review that has been done in terms of spatio temporal case. Here, discussed a motion detection approach to recognize human activities in the video sequences. Motion detection is extensively used in video surveillance application to detect and recognize the human and human activities. This paper presented various types of visual features and classifiers were discussed. They are considering spatio temporal feature points used for recognition simple and complex human activities. In Spatio temporal cases give the good results and to avoid in real time problems are namely illumination changes, moving background and shadow detection. Here, discussed and compared several classification methods for activity recognition in spatio and temporal case. This paper is related to KTH human activity dataset and used state-of-the art methods. Furthermore, proposes an adaptive based algorithm for computationally effective to detect and recognize the simple and complex human activities in real time video sequence.

Reference

[1] Joshua Candamo, Matthew, and Dmitry "A Survey on Human Behavior-Recognition Algorithms"IEEE Transactions on Intelligent Transportation Systems, vol. 11, no.1, march 2010.

[2] Enugu Kim, SumiHelal, and Diane Cook, "Human Activity Recognition and Pattern Discovery "IEEE Pervasive Computing, vol. 9, no. 1, pp. 48-53,January 2010.

[3] T. Guetal "epSICAR: An Emerging Patterns Based Approach to Sequential, Interleaved And Concurrent Activity Recognition"Proc. IEEE 7th Ann. Int'l Conf. Pervasive Computing and Comm. (PerCom 09), IEEECs Press, 2009 pp. 1–9.

[4] JieYang,Jian Cheng, and Hanqing Lu “Human Activity Recognition Based On the Blob Features” Proceedings of the 2009 IEEE international conference on Multimedia and Expo,September 2009.
[5] M. S. Ryooand , and J. K. Aggarwal , “Recognition of Repetitive Sequential Human Activity” Proceedings of the IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan,October2009.
[6] Z. Wang and B. Li, “Human Activity Encoding and Recognition Using Low-Level Visual Features,” in Proceedings of the 21th International Joint Conference on Artificial Intelligence, 2009.
[7] Thomas dean and Rich Washington,“Recursivesparse,Spatio-Temporal Coding”in Proceedings of the 11th International Symposium on Multimedia,2009.
[8] YuyaHanai, Jun Nishimura AndTadahiro Kuroda “Haar – Like Filtering for Human Activity Recognition Using 3d Accelerometer” Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th, 2009.
[9] Junsong Yuan and ZichengLiutechware: “Video-Based Human Action Detection Resources”, IEEE signal processing magazine September 2010.
[10] Jong T.Lee“Real Time Parking Detection In Outdoor Environments Using 1-D Transformation” IEEE Transactions on circuts and systems for video techmology, vol 19, no. 7, july 2009.
[11] Ashok veeraraghavan “Rate Invariant Recognition of Humans And Their Activities” IEEE Transactions on image processing, vol 18.no 6, june 2009.
[12] Mohammad helmi, S.M.T. Almodarresi “Human Activity Recognition Using a Fuzzy Inter Ferrence System” IEEE 2009, Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference onPages 20-24 Aug.2009.
[13] Jyh-yeongchang,”Fuzzy Rule Interference Based Human Activity Recognition”IEEE international symposium on intelligent control part of 2009 IEEE multi-conference on systems and control saint Petersburg, Russia,july 8-10,2009.
[14] J. Niebles, H. Wang, and L. Fei-Fei, “Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words,” International Journal of Computer Vision, vol. 79, no. 3, pp.299–318, 2008.
[15] Muhammad usmanghani khan and atifsaeed ”Human Detection In Videos”Journal of theoretical and applied information technology , 2005-2009 jatit.
[16] Macropedersoli, jordi Gonzalez,”Boosting Histrograms Of Gradients For Human Detection”department of informatica, computer vision center universitatautonoma de Barcelona,08193 bellaterra,spain.
[17] Ishrishabh“Human Detection In RGB Images” department of information and computer science university of California, Irvine.
[18] Q.Zhu, M. C. Yeh, K.T. Cheng, and S. Avidan. “Fast Human Detection Using Cascade Histograms of Oriented Gradients”, In CVPR’06:

Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition, pages 1491-1498, 2006.

[19] P. Doll´ar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior Recognition via Sparse Spatio-Temporal Features," in Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, October 2005, p. 65.

[20] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing Human Actions: A Local Svm Approach," in Proceedings of the International Conference on Pattern Recognition. IEEE Computer Society, 2004.