A SURVEY ON HUMAN ACTION RECOGNITION FROM VIDEO

[1] Megha Shah,[2] Nitin Pandya
Student, Assistant Professor
Department of Information Technology and Computer Engineering,
Shankersinh Vaghela Bapu Institute of Technology Gandhinagar, Gujarat, India

Abstract— Every moment counts in the recognition of the action a complete understanding of the human activity in the video requires labels according to the actions that occur, by placing several labels densely on a video sequence. Here are many issues related to the recognition of human action in videos, such as cluttered backgrounds, occlusions, variation in viewpoint, rate of execution, and camera movement. A large number of techniques in this search, the different datasets used are the KTH dataset, the Weizman dataset, the UCF Sport dataset, and so on. In this survey, we provide an overview of current methods based on the ability to solve these problems, as well as how these methods can be generalized and their ability to detect abnormal actions.

Key Words : Action recognition, KNN (K-Nearest Neighbors),SVM, feature extraction.

I. INTRODUCTION

The recognition of human action is an active subject in the field of computer vision. This is due The recognition of human action is an active subject in the field of computer vision. This is due in part to the rapid increase in the number of video recordings and the large number of potential applications based on automatic video analysis, such as visual surveillance, human-machine interfaces, sports video analysis and video recovery. Among these applications, one of the most interesting is the recognition of human action, in particular the recognition of high-level behavior. An action is a sequence of movements of the human body that can involve multiple parts of the body simultaneously. From the point of view of computer vision, action recognition consists of matching the observation (for example, video) with previously defined models and then assigning a label to it, that is to say a type of action. Depending on the complexity, human activities can be classified into four levels: gestures, actions, interactions, and group activities.

In recent years, much work has been done in various areas of computer vision research, such as video classification [1], resolution [2] and segmentation [3], and so on. However, research on the recognition of human activity on video has not been widely explored, due to the difficulties encountered in processing the temporal information of the video stream. Action recognition from a video stream can be defined as an automatic recognition of human actions using a pattern recognition system with minimum human-machine interaction. In general, an action recognition system analyzes certain video sequences or images to know the characteristics of a particular human action in the training process and uses the knowledge acquired to classify similar actions during the test phase. Among the first advanced approaches or recognition of human action, all these surveys use motion and texture descriptors calculated according to the spatio-temporal points of interest, built manually. in part, to the rapid increase in the number of video recordings and the large number of potential applications based on automatic video analysis, such as visual surveillance, human-machine interfaces, sports video analysis, and video recovery. videos. Among these applications, one of the most interesting is the recognition of human action, in particular the recognition of high-level behavior. An action is a sequence of movements of the human body that can involve multiple parts of the body simultaneously. From the point of view of computer vision, action recognition consists of matching the observation (for example, video) with previously defined models and then assigning a label to it, that is to say a type of action. Depending on the complexity, human activities can be classified into four levels: gestures, actions, interactions, and group activities.


(a) Actions: Running and Walking


(b) Interactions: Pickup phone call and Hugging

(c) Gestures: Waving and Bend

II. INTRODUCTION OF IMAGE PROCESSING

Intorduction of Image Processing

Image processing is a method of performing certain operations on an image to obtain an improved image or extracting useful information. This is a type of signal processing in which the input is an image and the output may be an image or features/characteristics associated with that image. Nowadays, image processing is one of the fastest growing technologies. It is also a critical area of research in engineering and computer science.

Types of Image Processing:

1-Digital image processing

In computer science, digital image processing consists of using computer algorithms to process digital images. As a subcategory or domain of digital signal processing, digital image processing has many advantages over analog image processing. This makes it possible to apply a much wider range of algorithms to the input data and avoid problems such as noise buildup and signal distortion during processing. Because images are defined in two (perhaps more) dimensions, digital image processing can be modeled as multidimensional systems.

2-Analog Image Processing

In computer science, analog image processing is an image processing task performed on two-dimensional analog signals by analog means (as opposed to digital image processing). In principle, all data can be represented under two types named 1.Analog 2.Digital if the pictorial representation of the data represented in analog wave forms can be named as an analog image. Eg. TV broadcasting in ancient times .. through satellite dish systems.While digital representation or data storage in digital form is called digital image processing, for example digital data. image stored in digital logic gates.

III REALTED WORK

In this section, the state of the arts in human action recognition are summarized. In recent decades, many different approaches have been proposed for detection, representation and recognition, and understanding of video events. Previous research on action recognition focused primarily on RGB videos, which gave a lot of feature extraction, action methods of representation and modelling.This author presented human detection and simultaneous behaviour recognition from RGB image sequences using the action representation method based on the application of the classification algorithm to the HOG descriptor sequence. human animated images. Others have used a hierarchical filtered motion (HFM) method to recognize human action in cluttered videos, as in Then applied a global spatial motion smoothing filter to MHI gradients to eliminate unreliable isolated movements or noisy. To characterize the spatial (appearance) and temporal (movement) characteristics used by HOG descriptor in the intensity image and MHI, respectively and the Gaussian Mixture Model (GMM) classifier for action recognition performance system. Many researchers have improved the performance recognition performance of RGBD data by calculating a local spatio-temporal characteristic from RGB data, skeletal joint functionality, and scatter data, coding feature combination methods. sparse as in. In, presented a comparison of several well-known pattern recognition techniques. they used movement history images (MHI) to describe these activities qualitatively and calculated Hu moments. And the system was tested with feature vectors extracted with support vector machines and K-Nearest Neighbour classifiers. Another method used for action recognition is based on the features acquired from 3D video data by applying the Independent Subspace Analysis (ISA) technique to the data collected by the RGBD cameras as in the following model. recognize the activities. The other researchers considered the human activity as composed of a set of sub-activities, that is, they calculated a set of features based on poses and movements, as well as information about image and clouds of points.

IV.ACTION RECOGNITION IN VIDEO

Humans recognize and easily identify video actions but automation of this procedure is difficult. Recognition of human action in video is of interest for applications such as automated surveillance, aging behavior monitoring, human-computer interaction, content-based video retrieval and video synthesis. In monitoring the activities of the daily life of the elderly, for example, the recognition of atomic actions such as "walking", "bending" and "falling" is essential for the analysis of activities. So far, we

have focused mainly on improving different components of a standard discriminatory upward framework (such as the widely used word bag approach (Fig.1)) for video action recognition.We have three main contributions on the detection of local salient motion characteristics, representation and classification of actions. Our contributions on a more descriptive representation of the action and a better classification are being revised and will be published here. Stay tuned.

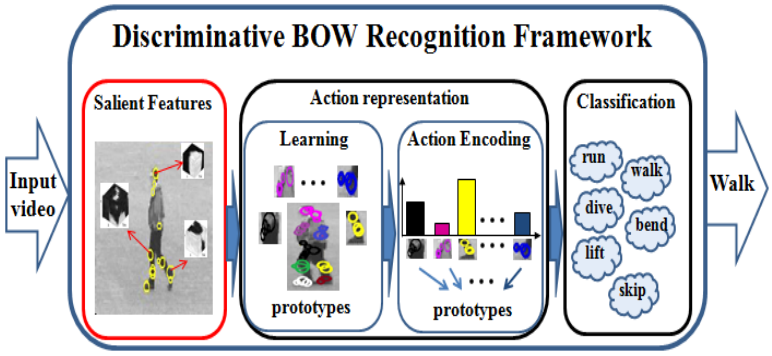

Fig.1: Standard Bag-of-Words framework for human action recognition.This section focuses on how to improve the detection of spatio-temporal salient features.

V.HUMAN ACTION RECONGITION

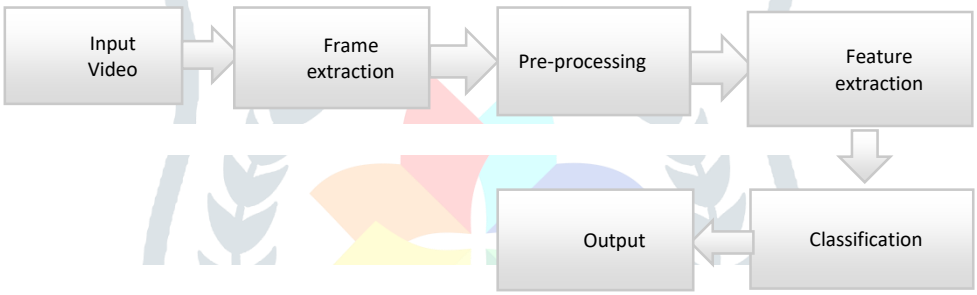

Fig.2: General Stage of Action Recognition

(a) Input Video: To recognize the human action video input is given to the system.

(b) Frame Extraction: The input video in which there are the activities of the human those frames are considered as a key frames and remaining frames are neglected.

(c) Pre-Processing: Pre-processing the image of abstraction whose purpose is to improve image data that suppresses unwanted deforms or enhances certain features of the image It does not increase the content of image information Its methods Neighboring pixels are the same or similar Brightness value and if a distorted pixel can be selected from the image, it can be restored as the average value of the neighboring pixels-the pre-processing in required to perform if the key frames contains illumination, shadow effects.

(d) Feature Extraction: Feature selection is one of the important parameter of any recognition system.

(e) Classification: The classification of images is one of the most complex areas of image processing. A machine learning technique is used assign a class tag to a set of unclassified data. in the classification techniques, there are two types of classification techniques, namely supervised Unsupervised classification and classification.

VI. TECHNIQUES USED FOR IMAGE CLASSIFICATION

1.SVM (Support Vector Machine):

The marginal sample is called a support vector machine. There are two types of SVM. First linear SVM where samples are separable linearly, which facilitates classification. For example, the 2-class classification and another non-linear SVM used to classify samples are not ranked in a linear fashion. For non-linear mapping, the samples are transferred to a larger dimension with the new dimension, so that the samples become linearly separable. To classify SVM multiclass, the binary classifier is converted to SVM multiclass. In the binary classifier, there are only two classes, which means that it is only used for two classes, and SVM multiclass can be used to classify more than two class labels. There are two approaches to multiclass SVM: one vs one and one vs all.

1.1. One vs One

In an approach vs one, each pair is separable. When two pairs of classes are considered, others are excluded. For the final result, the minimum distance of the generated vector has been calculated for the binary model representing each class.

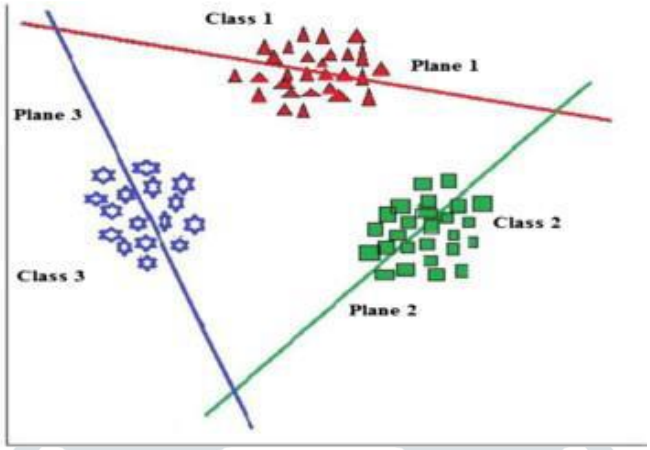

Fig. 3: Multiclass SVM

1.2.One vs All

The one against all approach constructs k distinct b classifiers k for the class classification k. The mth binary classifier is formed using the mth class data as positive examples and the k-1 remaining classes as negative examples. During the test, the class label is determined by the binary classifier which gives the maximum output value. Unbalanced learning is a major problem of the one-versus-rest approach. Assume that all classes have an equal size of learning examples, the ratio of positive examples to negative examples in each individual classifier is 1 k-1. In this case, the symmetry of the initial problem is lost. This involves dividing a N-class dataset into N two-class observations.

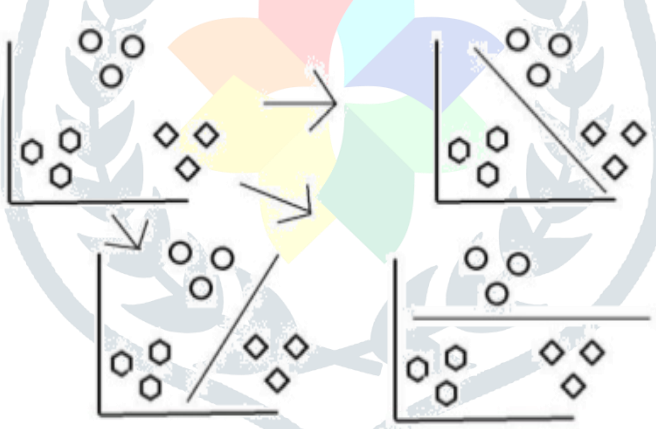

Fig.4: One vs ALL

2.K-Nearest Neighbors:

In the field of pattern recognition, K-Nearest Neighbor is one of the most important non-parametric algorithms. This is a supervised learning algorithm. The classification rules are generated by the learning samples themselves without any additional data. The K-Nearest Neighbor classification algorithm predicts the category of the test sample as a function of the K training samples, which are the closest neighbors of the test sample, and the judge according to the category having the highest category probability. great. The process of the K-Nearest Neighbor algorithm for classifying the sample X is as follows: Suppose there are j training categories C1, C2, ..., Cj and the sum of the training samples is N after the reduction of the characteristics they become a vector of the characteristics of dimension m. Make sure that the X sample is the same characteristic vector of the form (X1, X2, ..., Xm) as all the learning samples. Calculate the similarities between all the learning samples and X. By taking the ith sample di (di1, di2, ..., dim) as an example, the similarity SIM (X, di).

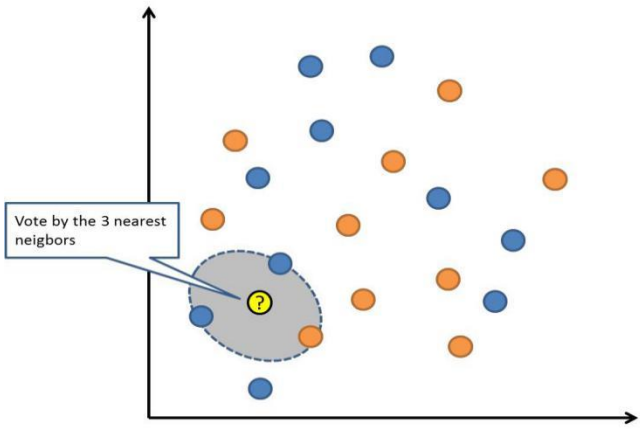

Fig.5: K-Nearest Neighbor

VII. CHALLENGES OF RECOGNIZING VISIONED ACTIONS

In this section, we list some of the challenges of research. facing the recognition of the action and describe the different methods used by researchers to manipulate them. A quantitative performance the comparison of the proposed techniques is difficult because The datasets and test strategy used vary considerably. However, the amount of training data and the ability to generalize method to any kind of action can be used as benchmarks to classify methods like these are essential to the success of the real world deployment [9].

A.Variation of point of view:
The movement patterns in each view path may appear specific, which makes the action's popularity less trivial. The most common technique for approaching the alternative in digital camera view attitude is to teach the classifier the use of multiple digital camera perspectives.

B.Occlusion:
Occlusions can be filled by objects in the camera's observation plane in anticipation of a video discovery or by self-occlusions.

C. Motion of the camera
The angle of the camera on which it captures video is the main expression of human recognition. The camera deals with the objection or the dynamics.

D .Cluttered Context
In video mixing, distraction is an advantage over a dynamic or cluttered background. It presents questionable information to identify the initial action.

E. Intra-class variation
Due to the lean modification within the class and the valuable inter-class variance, it is possible to control different human recognitions.

F.Appearance of the human
The human perception modifies what comes to the behavior of the actions of transmission which remain discrete according to the rise on which they are executed, bib and tucker also blew the cover of the role in the perception of the man and the objects they reinforce with them. This is why a search was made to remind the man to prepare a storm unrelated to the vision of the human being.

VIII. DATASETS

Testing the action recognition algorithm is essential as it provides qualitative and quantitative performance analysis, but for reliable analysis, it is necessary that the datasets capture all actions in the context of various challenges and conditions prove that the system is robust to them. Datasets that capture actions in all possible scenarios are very limited. In addition, some datasets are not publicly available. This is the main reason why many researchers create their own dataset for evaluation. A detailed study of the available data sets for performance evaluation is given in. The KTH dataset contains six action classes performed by 25 actors in four different scenarios: outside (s1), outside with scale (s2), outside with different clothes (s3) and inside (s4). The KTH dataset is small because the cameras are relatively stationary and only the zoom of the camera is considered a camera movement [9].

The Weizmann dataset contains ten share classes of nine actors using a simple background and a still camera. They also provide background sequences so that silhouettes can be extracted easily [9].

XI. CONCLUSION AND FUTURE WORK

In this proposed work, the combination of descriptors HOG and SURF gives a better discriminating power for the recognition of human actions. The experiments are performed on different datasets including KTH videos, Weizmann Dataset and also will be used Standard Datasets. the overall recognition rate is higher than that obtained with the HoG data set.

Future work will need to address many issues such as high computing costs, appearance changes due to clothing, changing lighting, and low recognition rates.

REFERENCES

[1] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," International Conference of Learning Representations, 2016.
[2] A. A. J. Johnson and F. Li, "Perceptual losses for real-time style transfer and super-resolution," in In European Conference on Computer Vision (ECCV), 2016.
[3] A. Kolesnikov and C. H. Lampert, "Seed, expand and constrain: Three principles for weakly-supervised image segmentation," in In European Conference on Computer Vision (ECCV), 2016.
[4] Y. Tian, L. Cao, Z. Liu, and Z. Zhang, "Hierarchical filtered motion for action recognition in crowded videos," IEEE Trans. Syst. Man Cybern.Part C Appl. Rev., vol. 42, no. 3, pp. 313–323, 2012.
[5] C. Thurau, "Behavior Histograms for Action Recognition and Human Detection," Hum. MotionUnderstanding, Model. Capture Animat., pp. 299–312, 2007.
[6] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human Activity Detection from RGBD Images," IEEE Int. Conf. Robot. Autom., pp. 842–849, 2011.
[7] N. Nguyen, "Feature Learning for Interaction Activity Recognition In RGBD Videos," CoRR, arXiv, pp. 2011–2013, 2015.
[8] Y. Lin, "Combining RGB and Depth Features for Action Recognition based on Sparse Representation," ICIMCS '15, August 19-21, 2015, Zhangjiajie, Hunan, China, no. 200, 2015.
[9] Manoj Ramanthan ,Wei-Yun, Yau EamKhwang Teoh Member of IEEE "Human Action Recognition With Video
[10] Human Action Recognition in Unconstrained videos by Explicit Motion Modelling, Yu-Gang Jiang, Qi Dai, Wei Liu, Xiangyang Xue, and Chong-Wah Ngo, IEEE
[11] Rawya Al-Akam,Dietrich Paulus "RGBD Human Action Recognition using Multi-Features Combination and K-Nearest Neighbours Classification" International Journal of Advanced Computer Science and Applications, 2017- IJACSA.