Flow trajectory approach for human action recognition

Trupti K Thakur[1], Prof. Ujwal Harode[2]

[1]*ME Student: Department of Electronics PIIT, New Panvel*
[2]*Professor: Department of Electronics PIIT, New Panvel*

---***---

Abstract - *Human action recognition and classification are very important for understanding an image. We propose algorithm for human action recognition with calculating different parameter to recognize action in effective way. Manually handling of video is very difficult .So we needed an automated analysis to process video which are collection of sequential images. In video analysis we need to follow steps like Detecting of human actions from video frames, tracking of that interested keypoints in consecutive frames, and Analysis of keypoints tracks to recognize their behaviour. Scale invariant feature transform method (SIFT) is method of video representation is used for extracting features frame to frame. It provide a convenient way for tracking and recognizing human action from video. In our proposed work scale invariant feature transform is used for feature extraction. Then frame by frame the feature which are extracted are tracked and parameter like magnitude and direction are calculated which used for recognizing motion.*

***Key Words*: Human action recognition, SIFT, Feature Extraction , Key points ,motion analysis.**

1. INTRODUCTION

Human action recognition is an important area of computer vision research and application. Here we are trying to recognize and track human action over a sequence of images. It provides information about the identity of a person and their personality which is difficult to extract. Human activity like walking, running are easily recognisable but activity like swing lifting are difficult to identify. Thus we needed an automated analysis rather than human operators monitored it. It aims to locating moving objects in a video file. The goal of the action recognition is an automated analysis of ongoing events and their context from video data. We have proposed an efficient algorithm analysis of the video frames and outputs the location of moving targets within the video frame. In the automated analysis we will extract feature first using sift algorithm and will apply flow trajectory approach frame to frame to recognise human action. The extracted features are tracked frame to frame .The tracked features can be analysed to recognise human action. Thus feature extraction from video frame acts as a first step for next processing such as tracking of the feature extracted frame to frame. The feature extraction plays a challenging task in real time application. Its application includes surveillance system, patient monitoring system, unmanned aerial vehicles, and sports play analysis a variety of system that involves interaction between person and electronic devices as human computer interfaces. Thus feature extraction play major role in various fields. In most of the application optical flow method is used.

Human action recognition can be applied to a wide range of fields nowadays such as multimedia, video data compression, industry production, and military affairs and so on. The feature extraction and tracking feature extracted in real time is very important task in image processing, computer vision.it combines different technologies as image processing, automation, information science and signal processing etc.

Various approaches are used by various researchers such as bag of visual word or bag of word approach. In such approach feature vectors are extracted from videos and this vectors from training dataset of same class are grouped to form clusters, these clusters are called as visual words. Two classification techniques are used in this approach as support vector machine and histogram matching respectively.

1.1 Overview of system

In proposed system the aim is to build robust and novel human action recognition, classification algorithm that can detect human in a variety of challenging real world scenarios. The Overall system overview will be represented in figure,

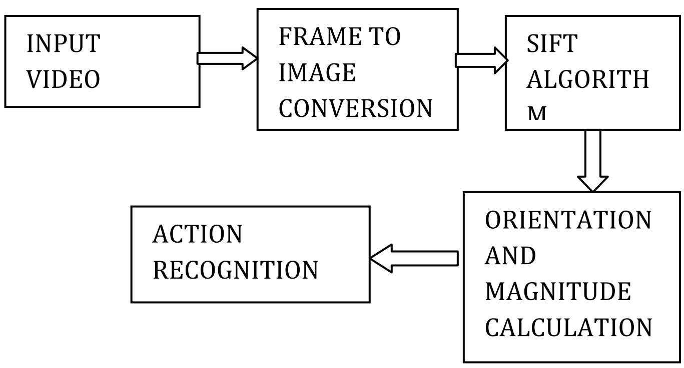

Fig-1: Proposed System

In proposed system the aim is to build robust and novel human action recognition, classification algorithm that can detect human in a variety of challenging real world scenarios. The Overall system overview will be represented in figure.

The steps involved in overall system are

1. The first step is the acquired video from database. The input video will be taken for both static and dynamic background.
2. For processing the video files, convert video into frames and from frames to images.
3. The next step is to apply SIFT.SIFT is used to extract the features of object and matching will helps to classify the objects in video frame.
4. Then next one is parameters orientation and magnitude are calculated for extracted feature keypoints.
5. Actions are recognised with the help of analysis of array for each game.

2. Experimental Result

The UCF Sports Action Dataset is a popular dataset to evaluate human action recognition algorithms. This dataset contains 35 videos from seven action classes, golf swinging, kicking, lifting, horse riding, walking, running, skating, swinging. These videos consist of realistic videos typically taken from broadcast television channels such as the BBC and ESPN. I use the same split of training/testing samples for our experiments, in which the dataset is divided into training and testing sets by taking two third of the videos from each action category to form the training dataset, and the rest one third of the videos are used for testing purposes.

Several experiments had been done to evaluate the feature extraction. These sequences used in experiments consist of golf swinging, kicking, lifting, horse riding, walking, running , skating and swinging videos so that the proposed scheme can be fully evaluated. First, target object of interest is defined from some frames. Then SIFT features are obtained from the video.

Frame Conversion:

For processing an Input Video files, it has to convert it into frames by finding the information about .avi file. After that it has to convert into images. So videos are split into frames. The first frame, which is called reference frame, which represents the reference pixel values for comparing purpose and the second frame which is called the input frame, which contains the moving object.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \text{ where} \quad (1)$$

$$G(x; y; \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-x^2+y^2}{\sigma^2}} \quad (2)$$

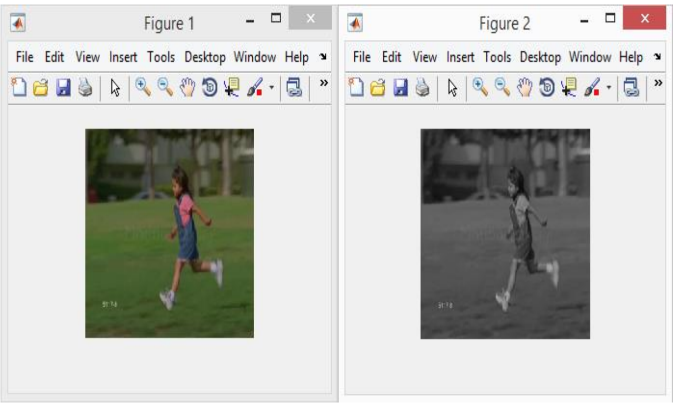

Fig-2: Human action video converted to initial image .grey level image.

Scale space extrema

Scale invariant interest points detected from a grey level image using scale space extrema of the Laplacian For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown in fig 5.2 on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated.

$$D(x; y; \sigma) = L(x; y; k\sigma) - L(x; y; \sigma) \quad (3)$$

This stage of SIFT is the detection of local interest points called keypoint. In this stage, the algorithm must search the potential keypoints over all scales and image locations. It can be efficiently implemented by using a difference-of-Gaussian function that are invariant to scale and orientation. The scale space of an image is defined as a function that is produced from the convolution of a variable-scale Gaussian G(x, y, σ), with an input image I(x, y).

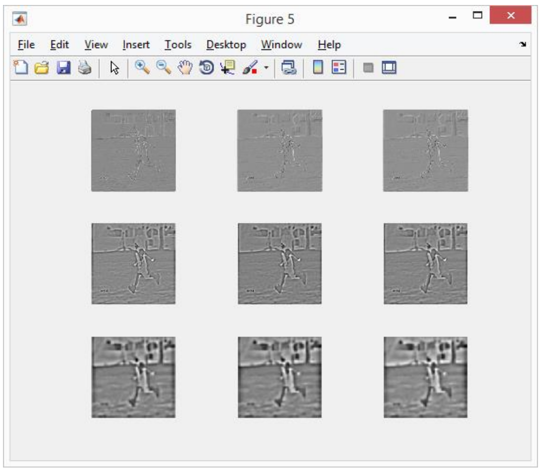

Fig -3: Scale invariant interest points detected from a grey level image using scale space extrema of the Laplacian For each octave of scale space,

Keypoint localization

To effectively detect stable keypoint location in scale space, Lowe used scale space peaks in Difference of Gaussian (DoG) function convolved with the image D(x,y,σ) which can be computed from the difference of two nearby scaled images separated by a multiplicative factor k.

The angle calculated using formula is in radiance so first it get converted into the degree form for further calculations .here we are considering 90° range of orientation bin size rather than considering 360° orientation bin size. If the vector in vector space rotated in any direction in 90° range of orientation we will get angle between 0° to 90°.Because of this complexity will reduce rather calculating for 360° orientation bin size.

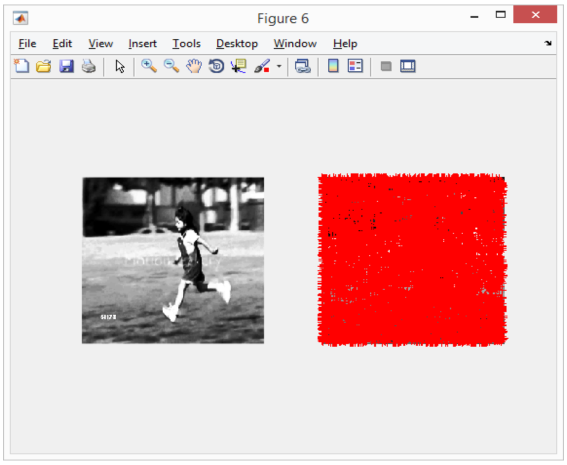

Fig-5: Keypoint Localization search each pixel in the DoG map to find the extreme point

The optical flow command in program is used for calculation amplitude and orientation for 15°.In analysis part we initially carried out experiments for various bin sizes out of which we get maximum clarity at bin sizes of 15° .we get clear distribution of orientation around the middle value. To check if the orientation of velocity vector for each pixel depending on in which bin it falls we increment that particular bin in array by one.by doing this we get to know no of orientation in particular bin.

Accurate keypoint localization

The next stage is to perform a detailed fit to the nearby data for location, edge response and peak magnitude. A location in image scale space is identified that are invariant with respect to image rotation, translation and scaling. At each candidate location, a detailed model is fit to determine location, scale and contrast. Keypoints are selected based on measures of their stability.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

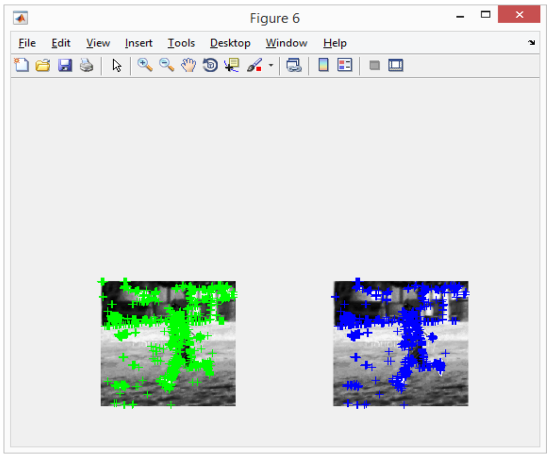

Fig-4 : Accurate keypoint localization by eliminating the point with low contrast or poorly localized on an edge.

Flow tracing

The idea is based on trajectory i.e. tracking of keypoints frame to frame. The keypoints are extracted by applying SIFT algorithm initial steps i.e. scale space extraction and key point localization as explained above.

Consider a keypoint $P_i^{(t)}$ in frame I_t then the tracked point $P_i^{(t+1)}$ in the next frame I_{t+1} is computed using median filter on keypoints. For each tracked point orientation and amplitude is calculated frame to frame for flow tracing of key points so the it will be easy to find the feature for that game.

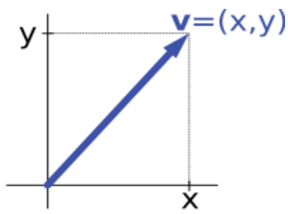

Fig-6 vector space

A trajectory is the path that a moving object follows through space as a function of time. A trajectory can be described mathematical either by the geometry of the path or as the position of the object over time. It will store the actual path of object of interest i.e. information of target in consecutive frames. We will get the all information about target object that in which direction it moves and what is the speed of target.

Then, SIFT features are obtained from the consecutive frames to match the feature from interested object. The features of frames are also stored by other keypoints descriptors.

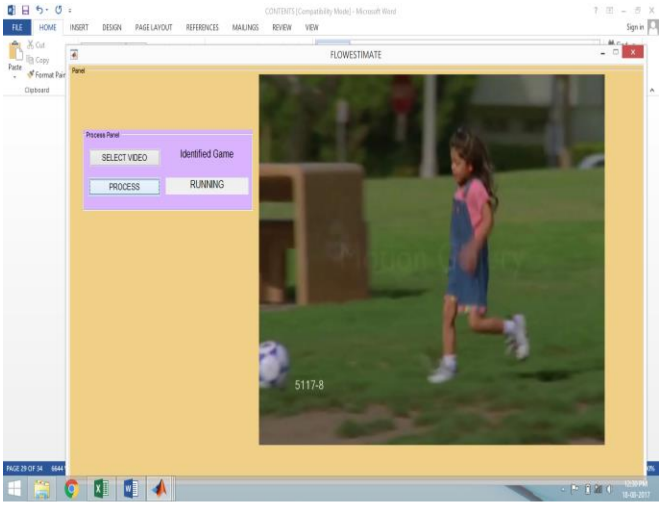

Fig-7: Human action recognized is displayed in identified game box

3. CONCLUSIONS

Human action recognition and tracking is an important task in computer vision field. In human action detection and tracking it consist of two major processes, human action detection and tracking Using SIFT feature extraction first feature of the object and the frame has detected to match the interested object. Since for feature extraction, SIFT algorithm has been used so tracker is invariant to representation of interested object.

REFERENCES

[1] J. K. Aggarwal and Q. Cai, Human motion analysis: A review, Computer Vision and Image Understanding, vol. 73(3), pp. 428-440, 1999.

[2] D. G. Lowe, Object Recognition from Local Scale-Invariant Features, International Conference on Computer Vision, 1999.

[3] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 2004.

[4] Babu.p Lakshmi, Devi.P,Object Detection and Classification Using Scale Invariant Feature Transform For SMS Reporting System International Journal & Magazine of Engineering, Technology, Management and Research , Vol(2), Issue No: 3 (March),ISSN No: 2348-4845,2015

[5] Video Object tracking using SIFT and Mean Shift, Chaoyang Zhu Department of Signals and Systems Signal Processing Group CHALMERS UNIVERSITY OF TECHNOLOGY Goteborg, Sweden, 2011

[6] M. D. Rodriguez, J. Ahmed, and M. Shah. Action MACH: A Spatiotemporal Maximum Average Correlation Height Filter for Action Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[7] C. Schuldt, I. Laptev, and B. Capito, Recognizing human actions: A local SVM approach, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3169 3176, Colordo Springs, United States, Jun. 2011