HAR Optical Flow Analysis In Static Background

Dr.S.Vimala[#1], P. Kalaivani[*2]
[#]*Associate Professor, Department of Computer Science,*
[*]*Ph.D. Scholar, Department of Computer Science,*
[#,*]*Mother Teresa Women's University, Kodaikanal, Tamil Nadu, India.*

Abstract -

Human Action Recognition (HAR) is a wide research area in computer vision. In this HAR model we have used the Weizmann dataset for training. In this the video is converted into a sequence of image frames. For training dataset we have stored only the first 25 image frames for feature extraction. Using this training dataset silhouette image features are stored and used for optical flow analysis. Background subtraction method is applied for extracting silhouette images from the Weizmann dataset. Gait function is used to analyse the HAR. We have compared our results with the Lucas-Kanade method for optical flow analysis, by reducing the matrix into 3-by-3 matrix. Our method reduces the noise level at a great extent and increases better optical flow vector value. Comparing the results with the existing method shows that we have attained better results.

Keywords - **HAR, training dataset, feature extraction, optical flow, silhouette.**

I. INTRODUCTION

In computer vision Human Action Recognition (HAR) has more attention in many applications such as medical and security management. Due to complex background settings, HAR is always hard when recognizing similar actions. In this paper, we explored novel human action recognition to learn contextual relationship between human actions and optical flow analysis in the static camera videos. We have used the dataset from Weizmann and UCF[1,2]. A generative probabilistic framework is used for action feature extraction from static background image scene. Some existing methods have results on a realistic video dataset validate the effectiveness of the human action recognition model for action from controlled background settings. Extensive experiments were conducted on different datasets and the Harris corner point[3,4] feature extraction method the learned model has good robustness when the features are noisy.

The contextual relationships of background settings could be learned as a complement for action recognition, as human actions always occur under particular action. The main problem of action recognition is based on methods how the background scenes and action categories for all videos were present. Otherwise, training dataset detectors performances of recognition will be affected by the learned detectors. In this paper, we intend to recognize the human actions in the training dataset[Weizmann] by using Gaussian gait function comparing with the test dataset[Weizmann, UCF]. The contextual relationship between actions and background scenes could be used to infer actions from background settings. By applying the gait function the actions are stored as training dataset. The training dataset has 6 type of actions from Weizmann[1] dataset actions of "Walk", "Run", "Jack", "Skip", "Jump" and "Bend". Firstly, a video will be divided into frames and blob frames of the silhouette are taken by subtracting the background regions. Then the features are extracted from the video frames by using the gait function. The actions are classified by using the SVM[5] classifier the proposed model gives better results for the training dataset. Finally the silhouette image frames are tracked for optical flow analysis.

II. RELATED WORKS

A. *Feature and Feature Detection*

A feature is defined as an "interesting" part of an image, and in many computer vision algorithms features are used as a starting point. The overall algorithm will only be valid and its feature detector works well based on the features used since they are the main primitives for the algorithms. Feature detection is a first level image processing operation, since it is usually performed as the first operation on an image. This detects that every pixel to check if there is a feature present at that pixel. The feature extraction algorithm[3,4,6] will typically only examine

the image in the region of the features. Gait function is used to extract the feature from the training image and is usually smoothed by a Gaussian kernel in a scale-space representation. Several feature images are computed, often expressed in terms of local image derivatives operations for train the image and for action recognition of the given image frames. A large number of feature detectors have been developed, since many computer vision algorithms use the feature detection as the initial step. To detect and characterize the periodic motion, they resort to Time-Frequency analysis. Parameswaran and Chellappa[7] propose a quasi-view-invariant approach, requiring at least five body points lying on a 3D plane or that the limbs trace a planar area during the course of an action.

B. Different Methods Based On The Image Features

1) Sobel and Canny Edge Detection Methods:

Edges are the points where there is a boundary between two image regions. An edge can be in arbitrary shape, and may include junctions. Edges are usually defined as sets of points in the image which have a strong gradient magnitude, some common algorithms(canny, sobel and so on) will then chain high gradient points together to form a more complete description of an edge in the image.

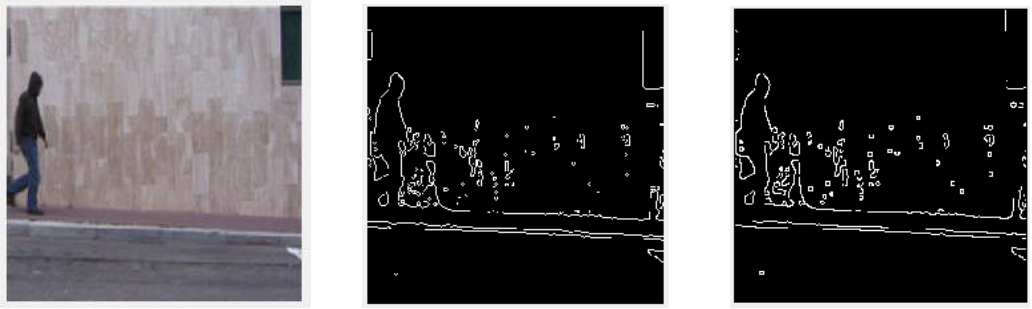

Fig.1 a) Walking action from Weizmann dataset applying b) Sobel's Method c) Canny's Method

Fig.1 shows the Walking human action image frames after applying Fig.1 a) Sobel's and b) Canny's edge detection[6] method for Weizmann dataset. These algorithms have some constraints based on the properties of an edge, like shape, smoothness, and gradient value. In general, edges have a one-dimensional structure.

2) Shi & Tomasi's and Harris Corner Detection:

The "corners" or "interest points" are the point-like features in an image, used for feature extraction. This algorithm finds rapid changes in corners. These corner detection algorithms such as Harris & Stephens, Plessey, Shi–Tomasi, etc.,[Fig.2] were developed for instance by looking for high levels of curvature in the image gradient.

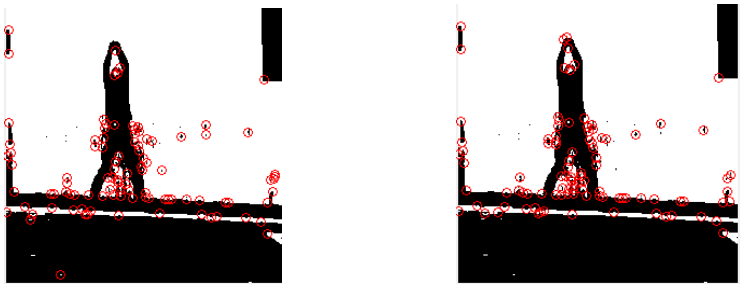

Fig.2 a) Shi & Tomasi's b) The Harris corner detector for the Human action (Jack action from Weizmann dataset)

The so-called corners were also being detected on the image in which the parts were not corners, that is, for instance on a dark background a small bright spot may be detected as an interest point. These points are known as interest points, but here we use the term corner.

3) Blobs Detection:

Blobs provide a complementary description of image structures in terms of regions that are different to corners. In contrast, the blob descriptors may often contain a preferred point such as the local maximum of an

operator response or a center of gravity which means that many blob detectors [Laplacian of Gaussian, MSER, PCBR, etc.,] may also be regarded as interest point operators.

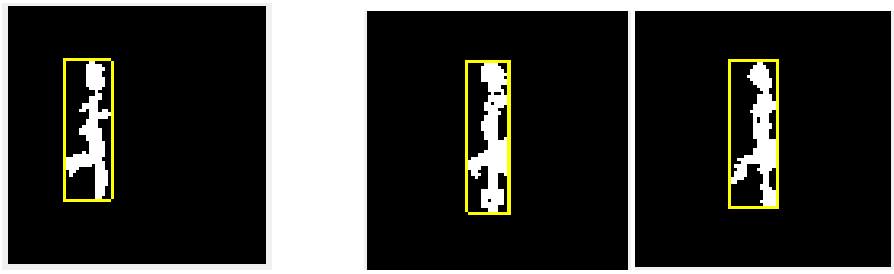

Fig.3 Blob frames for Human action tracking (Skip action from Weizmann dataset)

Blob detectors[8,9] works good for detecting the areas in an image which are too smooth to be detected by a corner detector. While shrinking an image and then performing corner detection, the detector will respond to points which are sharp in the shrunk image. Also corner detectors may be smooth in the original image. This was the main difference between a corner detector becomes somewhat vague than a blob detector. Here we have used the mean and deviation from the HSV-Gauss model with the data range from [0.25 0.6] by taking the threshold value as 6. The blob frames are shown in Fig.3. Here we have used the gait function for HAR by applying HSV Guass model to extract the features from the given frame.

4) *Ridges Detection:*

A ridge is a natural tool for the elongated object images. In general, ridge can be thought of as a one-dimensional curve that represents an axis of symmetry, and in addition has an attribute of local ridge width associated with each ridge point. It is algorithmically harder to extract ridge features from general classes of grey-level images than other feature detection methods. Nevertheless, ridge descriptors are frequently used in aerial images for road extraction and in medical images for extracting blood vessels.

C. *Local Features*

Local image features[3,4,6] which are also known as interest points, key points, and salient features[10,11] can be defined as a specific pattern that has unique from its immediately close pixels, which is generally associated with one or more of image properties. Such properties include edges, corners, regions, etc., as described. On the whole, a local feature typically has a spatial extent which is due to its local pixels neighborhood.

III. FEATURE EXTRACTION

After we have detected the features[12], a local image patch around the feature can be extracted. This extraction may involve quite considerable amounts of image processing. The result is known as a feature descriptor or feature vector. In addition to such attribute information, the feature detection[13,14] step by itself may also provide complementary attributes, such as the edge orientation and gradient magnitude in edge detection and the polarity and the strength of the blob in blob detection.

D. *Background subtraction*

Background subtraction[19] also known as foreground detection, which is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing for object recognition etc. Generally an image's regions of interest are objects like humans, cars, text etc. in its foreground.

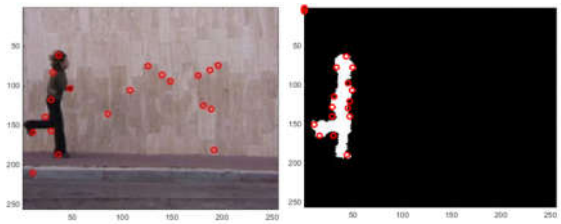

Fig.4 Harris feature detector a) Original image b) Background subtracted image (Skip Action from Weizmann Dataset)

Background subtraction is a widely used approach for detecting moving objects[9,15,16] in videos from static cameras. After the stage of image preprocessing (which may include image denoising, post processing like morphology etc.) object localization is required which may make use of this technique. Detecting the actions[9] from the video has difference between the current frame and a reference frame, often called "background image", or "background model". Background subtraction provides important cues for numerous applications in video surveillance (for example surveillance tracking or human poses estimation), human computer interaction, optical motion capture, content-based video coding, traffic monitoring, real-time motion gesture recognition.

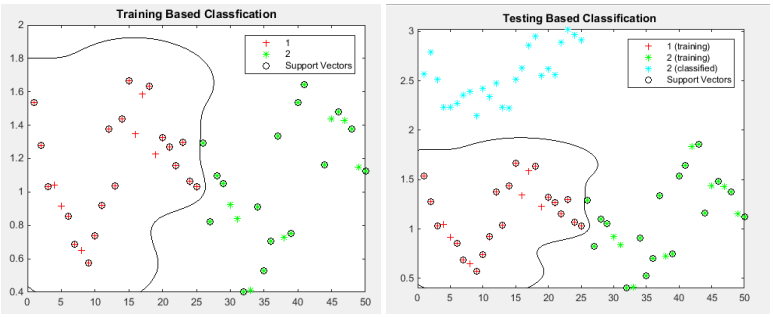

Fig.5 Results based on feature vectors using SVM classifier a) Training Dataset b) Testing Dataset

Background subtraction is generally based on a static background hypothesis which is often not applicable in real environments. With indoor scenes static backgrounds methods have difficulties, due to the reflections or animated images on screens lead to background changes. Similarly, with outdoor scenes, due to wind, rain or illumination changes brought by weather, static backgrounds methods have difficulties. The Fig.5 shows our experimental analysis result for two training set and one testing set. Here we have extracted the 25 frame features (Fig.4 shows the silhouette image feature extracted from the original image frame) from the video by applying Gauss model. C.W.Liang and C.F.Juang[18] proposes HOG feature in wavelet-transformed space with SVM classifier. Here we have used the SVM classifier[9] for Human action classification and the result shows that we have attained good results.

IV. FEATURE (HARRIS CORNERS) TRACKING

If we want to track one point in the object from the frame then use optical flow[19]. The motion of a point from one frame to another frame is called an optical flow.

Algorithm:

1. First we have to detect the Harris corner points in the image frame
2. Then calculate the Harris corner point, from the next consecutive frames
3. Detect the Harris corner points (the circle is used here)
4. Apply the gait function for silhouette images of the same image frame.
5. Harris corner points are tracked from the silhouette frame.
6. Then we have tracked the new and old Harris points.

In the proposed approach, the silhouette images are tracked by using the Harris detector for optical flow analysis, which reduces the noise level in greater level comparing to previous approach and are shown in the Fig.6.

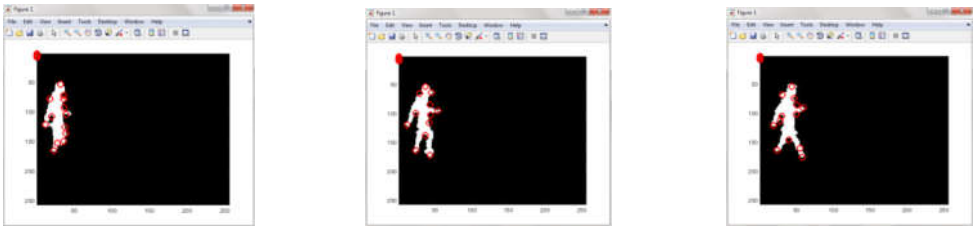

Fig.6 Harris corner points in the silhouette image

The Fig.6 explicitly illustrates how the proposed method is able to estimate the optical flow from human actions. This shows a person walking gestures with a feature vectors. The distinct interest points are detected at the moments and at the spatial positions where the structure of human changes its direction of the motion. We have to select the group of features from the given image that should be tracked in two subsequent frames to obtain the optical flow. We have used two different approaches to determine human actions. Lucas-Kanade algorithm is used for detecting the optical flow vector values from the given image frames.

V. EXPERIMENTAL RESULTS

E. LK Method For Optical Flow Analysis

The Lucas-Kanade method is widely used differential method for optical flow estimation developed by Bruce D. Lucas and Takeo Kanade. To compute the optical flow between two images, we must solve the following optical flow constraint equation.

$$I_x u + I_y v + I_t = 0$$

$$I_x u + I_y v = - I_t \qquad (1)$$

where I_x, I_y and I_t are the spatio-temporal[13] image brightness derivatives.

-u is horizontal optical flow

-v is vertical optical flow

It assumes that the flow is essentially constant in a local neighborhood of the pixel under consideration, and solves the basic optical flow equation for all the pixels in the neighborhood by the least square criterion. The optical flow methods try to calculate the motion between two image frames which are taken at times t and t+Δt at every voxel (ie, its position in the data structure that makes up a single volumetric image) position.

To solve the optical flow constraint equation for u and v, the Lucas-Kanade method divides the original image into smaller sections and assumes a constant velocity in each section. Then it performs a weighted least – square fit of the optical flow constraint equation to a constant model for $[u,v]^T$ in each section Ω. The method achieves this fit by minimizing the following equation

$$W^2 [I_x u + I_y v + I_t]$$

W is window function that emphasizes the constraints at the center of each section. The solution to the minimization problem is

$$\sum_{x \in \Omega} W^2 [I_x u + I_y v + I_t] \qquad (2)$$

LK Algorithm is a function which can be used to implement basic Lucas-Kanade algorithm[17] with only one level. These are defined as follows, two successive frames in a video are taken as input, and u,v are the estimated vectors after using Lucas-Kanade algorithm.

Set up the matrix of A by using the following function

$$A = [I_x \; I_y] \qquad (3)$$

Calculate the respective [u,v] at pixel (i,j) by using pseudo inverse

$$[u,v] = (A^T A)^{-1} A^T(-I_t) \qquad (4)$$

Our approach uses the (u,v) attained from the LK method, which is again iteratively processed to reduce the displacement vector points by reducing the window size into half. Also our method uses the 3-by-3 instead of 5-by-5 matrix of A to find the pseudo inverse which reduces the time and space vector.

Algorithm:

Step 1 : Get the two successive frames from the video(computed from Weizmann dataset)

Step 2 : Then estimate velocity vectors by using the Lucas-Kanade algorithm

Step 3 : Initialize the [u,v] vectors by zero.

Step 4 : Compute I_x, I_y and I_t values from the current consecutive two image frames.

for i = 1:size(image1)

I_x = image1;

for j = 1:size(image2)

I_y = image2;

Step 5 : Calculate the A matrix by using the I_x and I_y values, by using equation(3).

for (i=0 to 4)

compute $I_y = I_x[i]$ and $I_y = I_y[i]$; $I_t = I_t[i]$;

$A = [I_x \; I_y]$;

Step 6 : Calculate the respective [u,v] at pixel (i,j) by using the pseudo inverse.

$[u,v] = (A^T A)^{-1} A^T(-I_t)$, by using equation(4)

Step 7 : Compute the mean [u,v] function values.

Step 8 : Compute the iterative function of the A matrix to be reduced into 3-by-3.

(This could improve the precision of the optical flow tracking to a great extent).

Step 9 : This has 5-by-5 window size it could be reduced into 3-by-3, then current frame X has attained by

for i = 1 to iterations

[u, v] = LKIterate(u, v, Image1, Image2);

end

The previous [u,v] results are used to estimate and the current results that are based on this estimate and could be more precise than the previous ones. The optical flow experimental results have been shown in the Table I.

TABLE I

Dataset/ Algorithms	LK Algorithm	Our Approach
Walk	3.26E+04	5.67E+04
Run	4.37E+04	7.25E+04
Jump	3.38E+04	5.82E+04
Skip	4.86E+04	8.00E+04
Jack	3.50E+04	5.73E+04
Bend	3.26E+04	6.13E+04

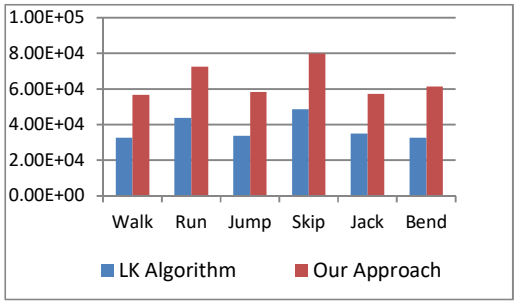

Fig.7 Optical flow analysis of HAR by using LK algorithm and our approach

The experimental results shows that we have attained better optical flow results comparing with LK method[Table I]. For walking action we have attained 57.49%, run 60.27%, jump 58.08%, skip 60.75%, jack 61.08% and bend 53.18% improved results compared with LK and is shown in graph[Fig. 7].

VI. CONCLUSIONS

In this paper, we explore to recognize human actions from background settings of controlled videos. Our proposed approach is used to analyze the human actions from the static view of background[9]. This method is used to infer human actions from dataset(Weizmann and UCF) according to a certain extracted features from the image. There are many ranges of potential applications available for the research of HAR. Experimental results validate that the addition of the silhouette images have much better results (ie, 58.48% in an average) for optical flow analysis and

improves the recognition precision. Besides, our proposed model has good robustness when applied in high resolution video datasets (UCF).

ACKNOWLEDGEMENT

The authors express their sincere thanks to Mother Teresa Women's University for conducting their research successfully.

REFERENCES

[1] *Fathi and G. Mori, "Action recognition by learning mid-level motion features", Conference on Computer Vision and Pattern Recognition (CVPR), 2008.*

[2] *M. 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.*

[3] *Chris Harris & Mike Stephens, "A Combined Corner And Edge Detector", Plessey Research Roke Manor, United Kingdom, The Plessey Company, 1988*

[4] *Wu Peng, Xu Hongling, Li Wenlin and Song Wenlong, "Harris Scale Invariant Corner Detection Algorithm Based on the Significant Region" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.3 (2016).*

[5] *M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. "Actions as space-time shapes", IEEE International Conference on Computer Vision (ICCV), 2005.*

[6] *Canny J F, "Finding Edges and Lines in Images", MIT technical report AI-TR-720,1983.*

[7] *Parameswaran.V and Chellappa.R, "View Invariance for Human Action Recognition," IJCV, vol. 66, no. 1, pp. 83-101, 2006.*

[8] *M. Ahmad and S. Lee, "HMM-based human action recognition using multiview image sequences," in Proc. ICPR, pp. I:263–266, 2006.*

[9] *P.Kalaivani and S. Vimala, "Human action recognition using background subtraction method", IRJET,2015.*

[10] *S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation" PAMI, 26(11):1475–1490, Nov 2004.*

[11] *M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, "Image Features Detection, Description and Matching", Springer Publications,2016*

[12] *Yali Amit, Donald Geman, and Kenneth Wilder, "Joint induction of shape features and tree classifiers", PAMI, 19(11):1300–1305, 1997.*

[13] *Muhammad Sharif, Muhammad Attique Khan, Tallha Akram, Muhammad Younus Javed, Tanzila Saba and Amjad Rehman, "A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection", EURASIP Journal on Image and Video Processing, 2017.*

[14] *S Maity, D Bhattacharjee, A Chakrabarti, "A novel approach for human action recognition from silhouette images", IETE J. Res. 63(2), 160–117 (2017).*

[15] *C. Stauffer, W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", Computer Vision and Pattern Recognition, Santa Barbara, CA, June 1998.*

[16] *Yifei Zhang, Wen Qu, Daling Wang, "Action-Scene Model for Human Action Recognition from Videos", 2nd AASRI Conference on Computational Intelligence and Bioinformatics, 2013.*

[17] *Lee I. B., Choi B. H., & Park K. S., "Robust measurement of ocular torsion using iterative Lucas-Kanade", Computer Methods and Programs in Biomedicine, 85(3), pp. 238-246, (2007).*

[18] *C-W Liang, C-F Juang, "Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers", Appl. Soft Comput. 28, 483–497 (2015).*

[19]K Suresh, "HOG-PCA descriptor with optical flow based human detection and tracking", Communications and Signal Processing (ICCSP), International Conference, IEEE, (2014).