Human Action Analysis With SVM Approach

Sushanta Reang
Instrumentation department
College of engineering pune
Pune, India
Susantameska2307@gmail.com

Dr S.L Patil
Instrumentation department
College of engineering pune
Pune, India
Slp.instru@coep.ac.in

Abstract—human action analysis in present days is a very common topic in computer based vision researches and other medical application. The main goal of human action analysis is to analysis of what events is going on from video data. In a video that is the input, local space time captures local events in terms of neighborhood points in video sequence and that can be changed accordingly in its size, velocity and the frequency of video patterns. A reliable system having feature of recognizing various human action has many applications in real time world. The few applications where action analysis is required include surveillance security systems, health-care monitoring, and many more systems that involve human computer interface devices. It fulfill our requirement as to how such variety of events can be utilize to analyze complex motion patterns also with various graph for various sequences of human action that can be captured from mobile. The video representation is constructed in terms of interest points and compared this representation with SVM classifier from given sets of data for recognition. The result shown for activity analysis shows the method of its advantage compared to other method for action recognition.

Keyword : SVM, Space scale, interest point, matching.

I. INTRODUCTION

Human activity analysis has provided enough interest in the many research areas because of its active use in various computing system. Many applications such as security system, retrieval of video, and user interface requires method for recognizing activity in different nature. Most difficult nature in video includes scene with blurred , changing background, camera is not stationary, different scale, action variation in movements and cloth of people they wear, change in different scale and interest point and further more. All of these condition introduce complications that have been solved in computer vision.

Recently, many method which successes in terms of recognizing and learning human action was developed. If we use measurement of image in terms of flow of optic, then the result of recognition of various features will depend upon the recording condition such as size of the image, occlusions and , movement of nonstationary camera.

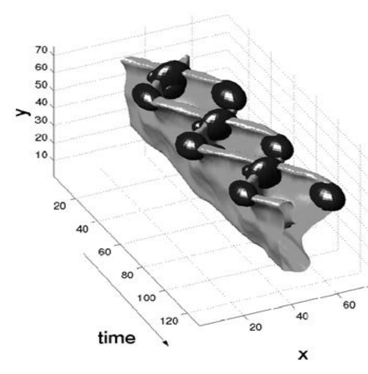
Moreover, huge measurement of image will be affected by motion of different objects and changes in cluttered frame. The solution to this problem is by means keeping the action movement stable or camera should be stable either, these solution however might be unstable in motion situation. This motivate the development of another solution that is video representation which is stable with respect to change of filming condition

So this paper explains, human activity analysis can be defined with local interest point in terms of spatio temporal point of interest.

More over huge data collected from mobile from different position of human action are compared with the real time action of stored data for further classification.

II. REPRESENTATION

To represent the motion in a video we have proposed LSTF (local space time feature) which here we considered as main events in resemble to moving in space time, two dimensional image structure at moment where image is of non constant (see figure 1).



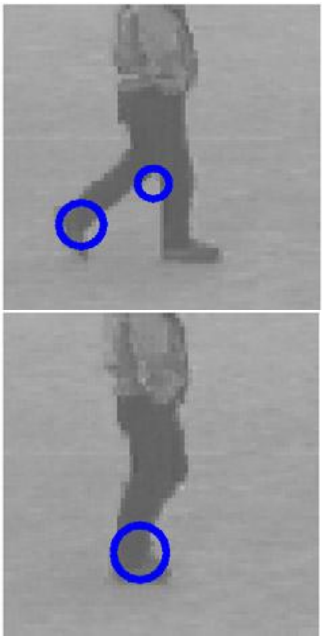


Fig a

Fig b

From the figure above, fig 1: LSTF for walking, fig a : leg motion for spatio temporal, fig b : events shown how the overlaying of video is, on selected frame.

Spatio temporal interest points in the field of spatial time can be expanded into local space time feature by getting the image value in spatio temporal to get huge change in both the space time and domain of temporal. The point will be spatial feature interest points having properties with a unique place in time dependence with moment with non stationary movement of image on a local space interest point neighborhood. In space time domain, the image can be modeled as $f^{sp} : R^2 \rightarrow R$ by its linear space scale representation. Now to detect the local feature in an image sequence f (x,y,t), we will construct its space scale representation $L(x, y, \sigma^2) = g(x, y, \sigma^2) * f(x, y)$ using Gaussian convolution kernel $g = \exp\left(-\frac{x^2+y^2}{2\sigma_l} -\right.$ t22τl2/√2π3πσl4τl2. The second moment matrix is computed using space time image $\nabla L = (L_x, L_y, L_t)^T$ with a Gaussian kernel convolution neighborhood of each point.

$$\mu(.; \sigma^2, \tau^2) = g(.; s\sigma^2, s\tau^2) * (\nabla L (\nabla L)^T) \qquad (1)$$

where '$*$' denotes the convolution operator, ∇L are the Gaussian derivative computed at local scale. Now, the Gaussian convolution can be used as covariance matrix of 2D presentation of image orientation in the event of space scale interest point. values of two large λ1, λ2 indicates the interest point is present in the neighborhood. To known exact such point, Stephen and harry have proposed to detect maxima of the corner by local maxima of

$$H = \mathrm{Det}\ (\mu) - \mathrm{ktrace}^3(\mu) \text{ over } (x, y, t).$$

The space scale interest point of an events in space scale and time is derived by spatial point and time scale parameter (σ, τ) of the Gaussian derivative kernel convolution. The size of an events can be adjusted accordingly to draw a match of interesting point of underlying image structures by selecting scale (σ, τ) . however, its size and shape of the feature can be matched to the frequency of space time events, hence, making the activity much more stable corresponding to the different amount of video movement. In this work we have used these method and developed features corresponding to scale and size to get invariance according to the shape of the moving events in the video and also as the relative velocity of the camera. Spatio temporal neighborhood of interest point contain data about the movement in a video and the human action of events in an video sequence. To collect this information, we have to compute space time neighborhood jets as shown

$$l = (l_x, l_y, l_t, l_{xx}, \dots \dots, , \dots .., l_{tttt}) \qquad (2)$$

To show points in a space time neighborhood we have taken space time Gaussian kernel derivatives defined $L_x m_y n_t^k = \sigma^{m+n}\tau^k(\partial_x m_y n_t^k g) * f$ now compute using selected scale values as (σ2, τ2), to compute the variance accordingly to relative camera movements, we need to wrap the neighborhood together using estimated velocity values prior to computation of 1. Now in order to compare two events in a video we need to compute the mahalanobis distance between their descriptor as $d^2(j1, j2) = (j1 - j2)\sum^{-1}(j1 - j2)^T$, here Σ is a matrix of covariance to the point of interest in the dataset. To compute such neighborhood points in video events, K-means clustering is applied in the spatio temporal point descriptors and groups of same points is detected with similar neighborhood.

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

III. CLASSIFIER : (SVM)

The classifier used here is SVM (support vector machine) which are state of the art with large margin classifier which got popular recently with visual features analysis. The separation is done in the set of training data into sets (X1, Y1), (X2, Y2), . . . (Xm, Ym) of two classes, assume that the classes of two sets can be divided by a hyperplane of w * x + b = 0 in space scale H. the values for b and w can be derived by using lagrange multipliers αi (i=1, . . . m).

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b\right)$$

α_i and b, the values of both can be known by using SVC learning algorithm.

IV. EXPERIMENT

SVM classifiers with movement classifier in terms of feature histogram (histLF) , local feature (LF) define two method for movement recognition. For evaluation, the input, a video dataset of six events of human action done many times by twenty five subjects in many different scenarios specially: done in outdoors with s1, with scale variation outdoors s2, wearing different clothes in outdoors s3 and indoors with s4, presently the dataset contains 2391 events. All the events were filmed over different backgrounds with a stationary camera with 25frame per second. And all the events were further classified according to the test subjects into a trained data. The input were first introduced on SVM training set while further validation set were done to categorize the parameters of each type of event. The final results of recognition were obtained on the test set. While on the other hand the mobile and the host computer are connected via internet ip address with certain id and password. With the help of mobile sensor as shown in figure (3) , that is the accelerometer the data for various sequences are transfer to the computer.

Figure 2: different sequences of events of walking, exercise, running, boxing, Hand wave, hand clap.

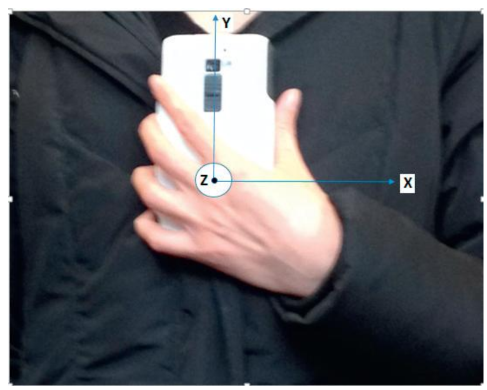


Figure 3: android sensor

V. RESULTS

To compute the action of various nature we have examined training on data different test subsets. It conclude that local features with support vector machine classifier gives the more best result to the sets of training , the examination of all results for different methods increases with data set used for training the sets. The figure shown below is tested on the training set of various scenarios. The frames of 50 is taken from each video for better evaluation. Depending upon the dataset it matches the classification with SVM approaches.

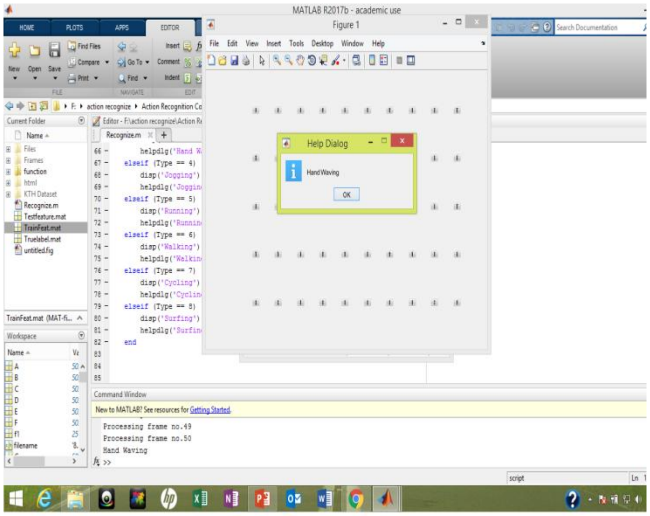
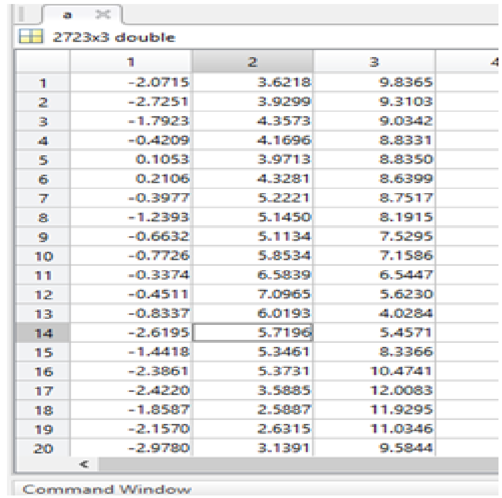


Figure 4: show the result for various sequences



a

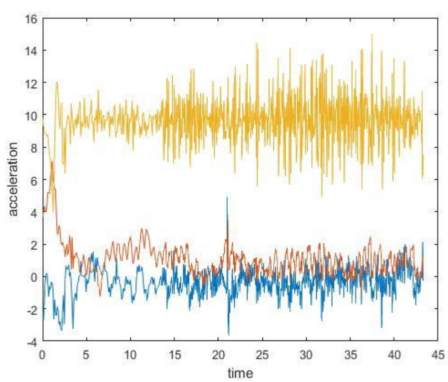
2723x3 double

	1	2	3	4
1	-2.0715	3.6218	9.8365	
2	-2.7251	3.9299	9.3103	
3	-1.7923	4.3573	9.0342	
4	-0.4209	4.1696	8.8331	
5	0.1053	3.9713	8.8350	
6	0.2106	4.3281	8.6399	
7	-0.3977	5.2221	8.7517	
8	-1.2393	5.1450	8.1915	
9	-0.6632	5.1134	7.5295	
10	-0.7726	5.8534	7.1586	
11	-0.3374	6.5839	6.5447	
12	-0.4511	7.0965	5.6230	
13	-0.8337	6.0193	4.0284	
14	-2.6195	5.7196	5.4571	
15	-1.4418	5.3461	8.3366	
16	-2.3861	5.3731	10.4741	
17	-2.4220	3.5885	12.0083	
18	-1.8587	2.5887	11.9295	
19	-2.1570	2.6315	11.0346	
20	-2.9780	3.1391	9.5844	

Command Window

Figure 5: result and observation from mobile sensor

Apart from the following, the data from the mobile is fetched to the computer and the graph for various activity. We have also evaluated the necessary of K value on the requirement of KNN clustered classification for each activity. By increasing the value of K it will effect the overall parameters of system, the best result can be obtained in the case where the selection of K is to done as 50, the window size is selected as 1.5 sec and interval of sampling is 50 m/sec. the figure (5) shows the result for accelerator.



VI. CONCLUSION

Thus the computation is done for how local space time events is utilize for recognition and representation motion patterns such as activity done by human. Adding together (LF) local feature with support vector machine classifier we have seen a method for activity analysis that gives high analysis results as compared to method of analysis.

Presentations of human activity patterns in terms of space time have great advantages of perfect and reliable to change in the movement, human variation, the action variation and background cluttered. We have also seen that space time feature give best recognition performance in scene with complex cluttered and unstable backgrounds. Whereas the performance with the data from mobile gives you the perfect representation of the action.

VII. REFERENCES

[1]. Q. Cai and J. Aggarwal, “human motion analysis: a review”. CVIU. 73-3: 428/440, 1999.

[2]. Laptev and T. Lindeberg, “space time interest point”. In proc. ICCV. Page 423-439, 2003.

[3]. Laptev, Christian Schuldt, Barbara Caputo, “recognizing human action: a local svm approach”. Computational vision and active perception laboratory.

[4]. Ravi, Nishkam, Dandekar, Nikhil, Mysore, Littman, Pretham, “activity recognition from accelerometer data, AAAI, 2005;5-1541.

[5]. N cistianini and J. Taylor, “ an introduction to support vector machine and other kernel based learning methods. Cambridge UP, 2000.
[6]. Krishnan, Narayanan, Colbry, Drik and Juillard, Colin, “ real time human activity recognition using tri axial accelerometers, sensors, signals and information processing workshop,2008.