International Journal of Computer Network and Security(IJCNS)
Vol 4. No 1. Jan-Mar 2012 ISSN: 0975-8283
www.ijcns.com

Multiview Unauthorized Human Action Recognition

P. Kalaivani[#1], **S.P.Audline Beena**[*2], **J. Rajalakshmi**[#3]
[#*]Department of Computer Science and Engineering,
Sri Muthukumaran Institute Of Technology, Chennai, India
[1]vijaykumarbtl@gmail.com
[2]audlinebeena@gmail.com
[3]jkrajee@gmail.com

Abstract

This paper presents the recognition of human actions under view changes. This deploys an automotive visual surveillance system to detect abnormal behavior patterns and recognize the normal ones. If a person enters a room, video of him is captured and stored(both front view and the top view) then it is given to the training module here the video is checked if it is a normal behavior splited image is taken, whenever the action is recognized blob images are saved, and the frame counts are taken. In case, the anomaly is detected the red color will be displayed. The abnormal behavior is achieved by keep tracking the videos and blob frames and checking each frame values. Most of the current methods for action recognition are designed for limited view variations. For a given action sequence and a given type of low level features, and compute distances between extracted features for all pairs of time frames and store results in a Self-Similarity Matrix (SSM). This project approach builds upon self-similarities of action sequences over time.

1. INTRODUCTION

Visual recognition and understanding of human actions have attracted much attention over the past three decades and remain an active research area of computer vision. A good solution to the problem holds a yet unexplored potential for many applications, such as the search for and the structuring of large video archives, video surveillance, human-computer interaction, gesture recognition, and video editing. Recent work has demonstrated the difficulty of the problem associated with the large variation of human action data due to the individual variations of people in expression, posture, motion, and clothing, perspective effects and camera motions, illumination variations, occlusions and disocclusion, and distracting effects of scenes surroundings. Also, actions frequently involve and depend on manipulated objects, which add another layer of variability. Most of the current methods for action recognition are designed for limited view variations.

A reliable and a generic action recognition system, however, have to be robust to camera parameters and different viewpoints while observing an action sequence. The multiview action recognition from a different perspective and avoids many assumptions of previous methods. Differently from the previous view-based methods, this does not assume multi view action samples either for training or for testing. This project approach builds upon self-similarities of action sequences over time.

2. RELATED WORK

The Multi view Unauthorized Human Action Recognition is based on the assumption that there exist well defined and known a priori classes of normal behavior patterns and using that find the abnormal behavior patterns. A silhouette is the image of a person, an object or scene consisting of the outline and a featureless interior, with the silhouetted object usually being black. From its original graphic meaning, the term "silhouette" has been extended to describe the sight or representation of a person, object or scene that is backlit, and appears dark against a lighter background. Anything that appears this way, for example, a figure standing backlit in a doorway, may be described as "in silhouette". Silhouette used in the fieldsoffashion and fitness to describe the shape of a person's body or the shape created by wearing clothing of a particular style or period.

Parameswaran and Chellappa[8]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. However, obtaining automatic and reliable point correspondences for daily video with natural human action is a very challenging and currently unsolved problem, which limits the application of the above mentioned methods in practice.

One alternative to the geometric approach is to represent the actions by samples recorded for the different views. A database of poses seen from multiple viewpoints has been created in Ahmadand Lee[1]. Extracted silhouettes from a test action are matched to this database to recognise the action beingperformed. The drawback of these methods is that each action needs to be represented by many training samples recorded for a large and representative set of views. Other methods perform a full 3D reconstruction from silhouettes seen from multiple deployed cameras. This approach requires a setup of multiple views, which again restricts the applicability of methods in practice.

One approach has a close relation to the notion of video self-similarity used byBenabdelkader, and Cutler and Davis [12]. In the domain of periodic motion detection, Cutler and Davis[2] track moving objects and extract silhouettes (or their bounding boxes). This is followed by building a 2D matrix for the given video sequence, where each entry of the matrix contains the absolute correlation between the two frames i and j. Their observation is that for a periodic motion, this similarity matrix will also be periodic. To detect and characterize the periodic motion, they resort to Time-Frequency analysis. None of the methods above explores the notion of self-similarity for multiview unauthorized action recognition.

Researchers have attempted to classify dynamic systems into different categories based on these textures. This concludes the proposed SSMs. As we shall see, SSMs are not strictly invariant under projective or affine transformations, but are experimentally found stable under 3D view changes.

3. SELF SIMILARITY MATRIX

TheSelf-similarity matrices have already appeared in the past under various specific forms, including binary recurrence plots associated to time series. For a sequence of images $I = \{I_1, I_2, \ldots I_T\}$ in discrete (x,y,t) space a SSM of I is a square matrix of size T x T,

$$[d_{ij}]_{i,j=1,2\ldots T} = \begin{bmatrix} \mathbf{0} & d_{12} & d_{13}\ldots & d_{1T} \\ d_{21} & \mathbf{0} & d_{23}\ldots d_{2T} & \\ . & . & . & . \\ . & . & . & . \\ . & . & . & \\ d_{T1} & d_{T2} & d_{T3}\ldots\mathbf{0} & \end{bmatrix} \qquad (1)$$

wheredij is the distance between certain low level features extracted in frames Ii and Ij, respectively. The diagonal corresponds to comparing a frame to itself(no dissimilarity) and hence is composed of zeros. The exact structures or the patterns of this matrix depend on the features and the distance measure used for computing the entries dij. The computed matrix patterns have a significant meaning for their application the diagonals in the matrix indicate periodicity of the motion.

For example, after tracking walking people in a video sequence, Benabdelkader, and Cutler and Davis[12] compute a particular instance of SSM where dij is the absolute correlation between two frames, as depicted in Fig. 1.The computed matrix patterns have a significant meaning for their application the diagonals in the matrix indicate periodicity of the motion.

For an action of opening a cabinet door, performed by two different actors from considerably different viewpoints, these points are depicted in Fig. 1a. Figs. 1c and 1e shows the SSMs computed for these two actions based only on one hand trajectory, where red color indicates higher values and dark blue color indicates lower values. The dynamic instances, red stars in Fig 1b and Fig 1d, correspond to valleys of different area/spread in this plot of SSM marked by magenta circles along the diagonal of the matrix. The exact spread of these valleys depends on the width of the peaks in the spatiotemporal curvature of the actions, as shown in Figs. 1b and 1d. However, whereas Rao capture only the local discontinuities in the spatiotemporal curvature, the SSM captures more information about other dynamics of the actions present in the off-diagonal parts of the matrix. Note also that the proposed notion of self-similarity, unlike does not require estimation of point correspondences or time-alignment between different actions.

In Fig.1.the relationship between proposed SSM representation and dynamic instances introduced. Two actors perform the action of opening a cabinet door, where the hand trajectory is shown in (b) and (d). The SSMs computed for these two actions based only on one hand trajectory are shown in (c) and (e), respectively. The "dynamic instances", marked in red stars in (b) and (d), represent valleys in the corresponding SSM, depicted by magenta circles in (c) and (e), respectively. The spread of each valley depends on the peak-width of the corresponding dynamic instance.

A. Trajectory-Based Self-Similarities

In trajectory based SSM, if a set of M points x^m, m=1…M, distributed over a person is "tracked" (in a sense to be specified later) over the duration of an action performance, the mean euclidean distance between each of the k pairs of corresponding points at any two instants i and j of the sequence can be computed as

$$d_{ij} = 1/M \sum \| x_i^m - x_j^m \|_2 \qquad (2)$$

where m=1...M X_i and Xj indicate positions of points on the track k at time instants i and j. The self-similarity matrix computed from (1) by SSM-pos.

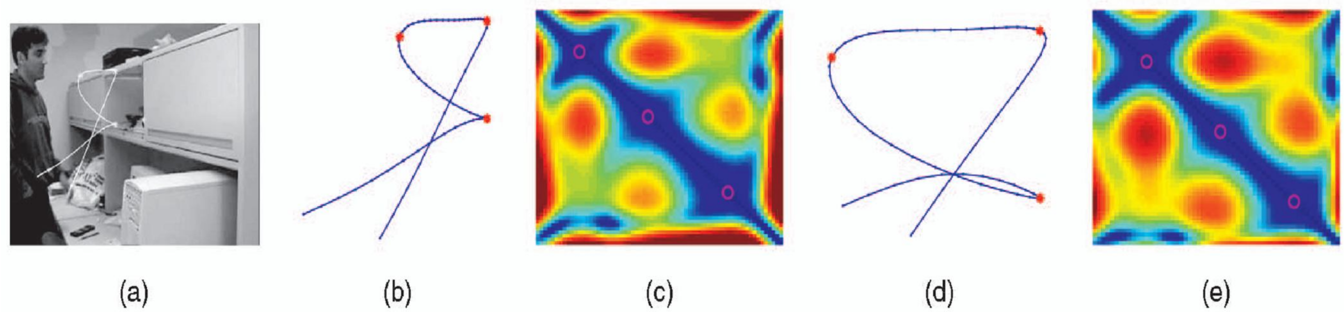

Figure1.Relationship between SSM

The overall goal of the proposed work being the recognition of actions in videos irrespective to viewpoints, the actual computation of SSM-pos requires that points be extracted and tracked in the input video. Assuming that this task is handled automatically by an external module, point tracker. Note that this method is not restricted to any particular subset of points as far as the points are distributed over moving body part

The definition of SSM-pos in (3) needs, however, to be adapted to a set of tracks with arbitrary length and starting time

$d_{ij} = 1/ |s_{ij}| \sum \| x_i^m - x_j^m \|_2 ,$ (3)

where m belongs to S_{ij} and $S_{ij} = \{1....M\}$ is the set with indices of point trajectories that are alive between frames i and j.

B.Image Based Self Similarities

Besides point trajectories, alternative image features can be used to construct other SSMs for the same image sequence. To describe the spatial appearance of a person at each image frame, compute Histograms of Oriented Gradients features. This descriptor, originally used to perform human detection, characterizes the local shape by capturing the gradient structure. In implementation, use four bin histograms for each of 5x7 blocks defined on a bounding box around the person in each frame.

Feature distance dij between time instants i and j is then computed as the euclidean distance between two HoG vectors extracted from frames Ii and Ij. SSMs computed using HoG features denoted by SSM-hog. In addition to HoG features, test the proposed method by considering optical flow vectors as another input feature. The corresponding SSMs are denoted by SSM-of.

	bend	jack	jump	pjump	run	side	skip	walk	wave
bend	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jack	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jump	0.0	0.0	77.8	0.0	0.0	0.0	11.1	11.1	0.0
pjump	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
run	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
side	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
skip	0.0	0.0	20.0	0.0	10.0	0.0	70.0	0.0	0.0
walk	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
wave	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

(a)

	bend	jack	jump	pjump	run	side	skip	walk	wave
bend	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jack	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jump	0.0	0.0	66.7	11.1	0.0	11.1	11.1	0.0	0.0
pjump	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
run	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
side	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
skip	0.0	0.0	12.5	0.0	0.0	0.0	87.5	0.0	0.0
walk	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
wave	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

(b)

Figure. 2. SSM based on action recognition on the Weizman single view action data set (Top)Example of nine classes of actions. (Bottom) (a) Image BasedSelf similarity SSM-of and (b) Trajectory Based Self similarity SSM-pos.

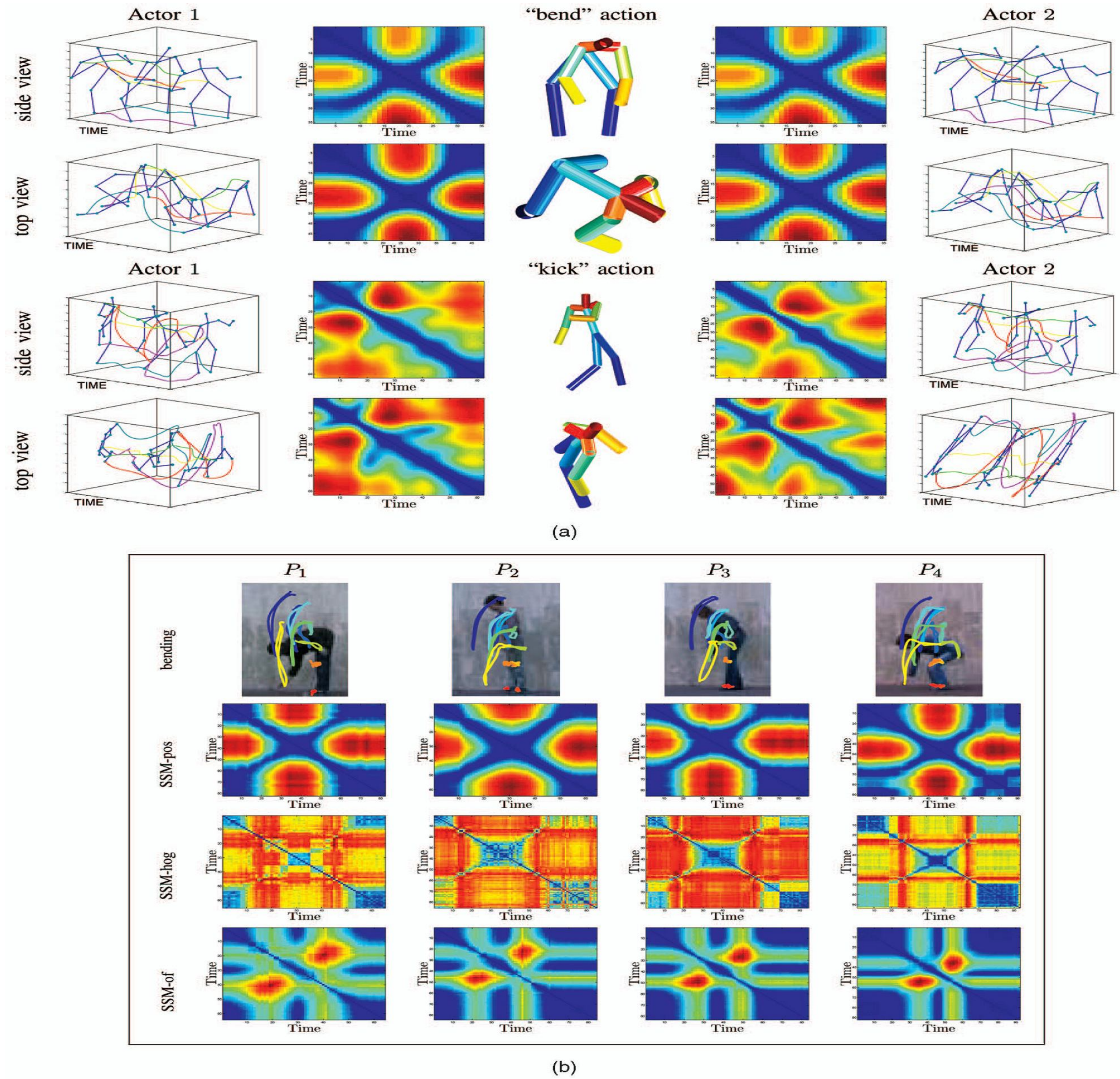

Figure. 3. SSM of different types

C.SSM-Based Action Recognition

SSM based action recognition, in this evaluate SSM-based video descriptors for the task of view-invariant action recognition. To recognize action sequences, follow recently successful bag-of features (BoF) approaches and represent each video as a set of quantized local SSM descriptors with their temporal positioning in the sequence being discarded.

In Fig.3 SSMs for different types of features and for different actions are given. (a) Examples fromthe CMU mocap data set.Columns 1 and 5 represent two actors, while columns 2 and 4 represent corresponding SSM-pos computed with 13 projected point trajectories, respectively. Different rows represent different actions and viewing angles. Note the stability of SSMs over different views and people performing the same action. (b) Examples from the Weizman video data set. Row 1: four bending actions along with manually extracted point trajectories used for computing SSM-pos; rows 2, 3, and 4 represent SSM-pos, SSM-hog, and SSM-of, respectively, forthese four bending actions. Note the similarity column wise.

Taking this view that global temporal ordering is not taken into action (as opposed to its use for synchronization where it is crucial information) permits to filter out fluctuations between actions from the same class while

D. Video Segmentation

Automatically segment a continuous video sequence ‘V’ into ‘N’ segments

$$V = \{V_1, V_2, V_3, \ldots\ldots Vn\},$$

where each segment contains a single behavior pattern. Each video segment Vn consists of Tn image frames like $Vn=\{Im_1, Im_2, Im_3 \ldots. Im_{Tn}\}$. The video can be simply sliced into overlapping segments with fixed time duration. In each splited image frames the blob frames are taken for behavior profiling. Both front and top view blob images are taken and stored for behavior profiling.

E .Event Based Behavior Representation

Identify the foreground and background pixel of a frame. Background model stores the values of a particular pixel which corresponds to the background colors. Pixel Change History (PCH) is represented for a pixel. Similar foreground pixels are grouped to form a blob. A behavior pattern is represented as a sequence of various events. For example, behavior patterns A and B contains events of classes a,b and a,c and e. Behavior patterns A and B are deemed as different since the events and their orders differs.

Build training data set and group training behavior patterns upon which a model for normal behavior can be built.

F. Anomaly Detection

The video to be detected is divided into segment. Each segment is divided into frames. Event is detected for each frame. The event is compared with the behavior model. If event not exists already in training video, it is considered as anomaly.

4. CONCLUSION

In this paper, we specifically addressed for multiview (front and top) independent video analysis, with human action recognition as a central application. Self-similarity being possibly defined over a variety of image features, either static (histograms of intensity gradient directions) or dynamic (optical flows or point trajectories), these descriptors can take different form and can be combined for increased descriptive power. Experimental validation on action recognition, as well as for the different problem of action synchronization, clearly confirms the stability of this type of description with respect to view variations. Results on public multiview action recognition data sets demonstrate superior performance of this method compared to alternative methods in the literature. Such encouraging results are simply obtained by exploiting the stability across views of SSM patterns.This method only makes mild

retaining sufficient action discrimination to build a view-independent action recognition system.

assumptions about the rough localization of a person in the frame. This lack of strong assumptions is likely to make this approach applicable to action recognition beyond controlled data sets when combined with modern techniques for person detection and tracking.

Acknowledgement

The Authors expresses their sincere thanks to the Department of Computer Science and Engineering, Sri Muthukumaran Institute of Technology Chennai for providing necessary facility to conduct the research work.

REFERENCES

[1]Ahmad.M and Lee.S, “HMM-Based Human Action Recognition Using Multiview Image Sequences,” vol. 1, pp. 263-266,2006.

[2]Cutler.R and Davis. L, “Robust Real-Time Periodic Motion Detection, Analysis, and Applications,” vol. 22, no. 8, pp. 781-796,2000.

[3]Dalal.N and Triggs.B, “Histograms of Oriented Gradients for Human Detection,” vol. 1, pp. 886-893,2005.

[4]Imran N Juneo, Emilie Dexter,Ivan Laptep and Patrick Perez, “View Independent Action Recognition From Temporal Self Similarities,” vol 33, no1, pp 172-185,2011.

[5]Iwanski.J and Bradley.E, “Recurrence Plots of Experimental Data: To Embed or Not to Embed,” vol. 8, no. 4, pp. 861-871,1998.

[6]McGuire.G, Azar.N , and Shelhamer.M, “Recurrence Matrices and the Preservation of Dynamical Properties,” vol. 237, nos. 1-2, pp. 43-47,1997.

[7]Moeslund.T , Hilton.A, and Kruger.V, “A Survey of Advances in Vision-Based Human Motion Capture and Analysis,” vol. 103, nos. 2-3, pp. 90-126,2006.

[8]Parameswaran.V and Chellappa.R, “View Invariance for Human Action Recognition,” vol. 66, no. 1, pp. 83-101,2006.

[9]Wang.L, Hu.W, and Tan.T, “Recent Developments in Human Motion Analysis,” vol. 36, no. 3, pp. 585-601,2003.

[10]Natarajan and R. Nevatia, “View and Scale Invariant Action Recognition Using Multiview Shape-Flow Models,” Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.

[11]Yilmaz.A and Shah.M, Actions Sketch: A Novel Action Representation, vol. 1, pp. 984-989,2005.

[12]C. Benabdelkader, R. Cutler, and L.Davis, “Gait Recognition Using Image Self-Similarity,” EURASIP J. Applied Signal Processing, vol. 2004, no. 1, pp.572-585, Jan, 2004.