Self-Similarity Estimation for Action Recognition using Gaussian Classifier

S.Bamini ,
PG Student
Department of ECE
Satyam College of Engineering
E-mail : baminisam@gmail.com

Daphni S
Research Scholar
Department of ECE
N.I University
E-mail : daphnithavasumony@gmailcom

Abstract- Action recognition is an active area of research and it attracted much attention from the researchers over the years .Appearance based methods are prone to differences in appearance between the training dataset and the testing sequences. In this project the dynamics of an action in video data forms a sparse self-similar manifold in the space-time volume, which can be fully characterized by a linear rank decomposition. Inspired by the recurrence plot theory, it introduce the concept of Joint Self-Similarity Volume (Joint-SSV) to model this sparse action manifold, and hence propose anew optimized rank-1 tensor approximation of the Joint-SSV to obtain compact low-dimensional descriptors that very accurately characterize an action in a video sequence. Our experimental results on five public data sets demonstrate that our method produces promising results and out performs many baseline methods.

***Key words*: Action recognition, self-similarity, tensor approximation.**

I INTRODUCTION

Action recognition has continued to be an active area of research and has thus rightfully attracted much attention from the researchers over the years. Important application domains, such as automatic video indexing and archiving, video surveillance, human-computer interaction, augmented reality, user interface design, and human factors would benefit immensely from a robust and efficient solution to this problem. as well as the robustness of the algorithm being used.

There are many factors that make this a challenging problem, including the large variations in performing an action by different people, whether by varying the postures, other execution speed, illumination variations in the sequences, occlusions and disocclusions, distracting background motions and perspective effects and camera motion. On the basis of representation, they can be categorized as: time evolution of human silhouettes, space-time shapes, dense trajectories, and local 3D patch analysis, generally coupled with some machine learning techniques.

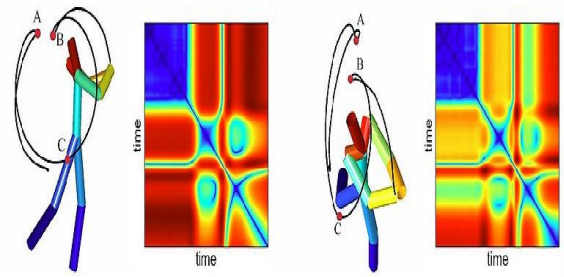


Figure 1.1: Similarity plot, a variant of recurrence plot, obtained for different views of human actions are shown to produce similar patterns

Joint Self-Similarity Volume (Joint SSV). Joint SSV is then decomposed into its rank-1 approximation vectors using an optimized iterative tensor decomposition algorithm. This yields a set of compact vector descriptors that are highly discriminative between different actions. To evaluate our method

on human action recognition, we used three public datasets. To show that our method is generic and does not depend on the input feature vector, we tested our method using low-level features like silhouette, as well as middle-level features like HOG3D. The final step used a nearest neighbor classification using the descriptor vectors produced by the rank-1 decomposition of Joint SSV.

1.1 Joint Self-similarity volume

Definition 1: An SSM can be expressed by a N×matrix Ri,j (η,v)=Θ(ηvi −vjp),i,j =1, ..., N, where N is the length of a feature vector v, and η is a threshold distance. The threshold η filters the values of each SSM element. We set η =0in this paper because this will give us a complete representation for the Joint SSMs. Θ(·)can be the Heaviside function (i.e. θ (x)=0,if x<0, and θ(x)=1otherwise) and · is chosen as an p-norm in this paper.

It can be verified that the SSM holds the following properties: Ri,j =Rj,i (Symmetry); Ri,j >=0for alli and j (Positivity); and Ri,k <=Ri,j +Rj,k for all i, j, k(Triangle inequality), and hence it is a metric. SSM provides important insights into the dynamics of a vector, which is especially advantageous in high dimensional spaces. The intuition behind the SSM is that, according to recurrent plot theory, if we view the vector v as a trajectory in 2D space, the SSM itself captures the internal dynamics of this trajectory in a matrix form.

Definition 2: The Joint SSM is defined as JRv, w i,j (ηv,ηw,v,w)=Θ (ηvvi−vj p1) Θ (ηw−wi−wjp2), in which i, j =1, ..., N, ηv and ηw are two internal thresholds,p1andp2are two distance norms.

This extension is motivated by the fact that a recurrence will take place if a point vj on the first trajectory v returns to the neighborhood of a former point vi, and simultaneously a point wj on the second trajectory w returns to the neighborhood of a former point wi

2 LITERATURE SURVEY

Free Viewpoint Action Recognition Using Motion History Volumes

Action recognition is an important and challenging topic in computer vision, with many important applications including video surveillance, automated cinematography and understanding of social interaction. Yet, most current work in gesture or action interpretation remains rooted in view-dependent representations. This paper introduces Motion History Volumes (MHV) as a free-viewpoint representation for human actions in the case of multiple calibrated, and background-subtracted, video cameras. This method present algorithms for computing, aligning and comparing MHVs of different actions performed by different people in a variety of viewpoints. Alignment and comparisons are performed efficiently using Fthisier transforms in cylindrical co-ordinates around the vertical axis. Results indicate that this representation can be used to learn and recognize basic human action classes, independently of gender, body size and viewpoint.

3d Shape Context And Distance Transform For Action Recognition

This method propose the use of 3D (2D+time) Shape Context to recognize the spatial and temporal details inherent in human actions. This method represent an action in a video sequence by a 3D point cloud extracted by sampling 2D silhouettes over time. A non-uniform sampling method is introduced that gives preference to fast moving body parts using a Euclidean 3D Distance Transform. Actions are then classified by matching the extracted point clouds. This proposed approach is based on a global matching and does not require specific training to learn the model. This method test the approach thoroughly on two publicly available datasets and compare to several state-of-the-art methods. The achieved classification accuracy is on par with or superior to the best results reported to date.

This 3D-Shape Context is an extension of the 2D Shape Context proposed by Belongie et al. [1] by including the temporal dimension. It differs from Kortgen et al.'s [11] 3D Shape Context, which does not extend an important property of the 2D shape context to 3D: Bins of equal distance from the origin should have the same size.

Behavior Recognition Via Sparse Spatio-Temporal Features

A common trend in object recognition is to detect and leverage the use of sparse, informative feature points. The use of such features makes the problem more manageable while providing increased robustness to noise and pose variation. In this work develop an extension of these ideas to the spatio-temporal case. For this purpose, this method show that the direct 3D counterparts to commonly used 2D interest point detectors are inadequate, and this method propose an alternative. Anchoring off of these interest points, this method devise a recognition algorithm based on spatio-temporally windowed data. This method present recognition results on a variety of datasets including both human and rodent behavior.

A new spatio-temporal interest point detector was presented, and a number of cuboid descriptors were analyzed. This method showed how the use of cuboid prototypes gave rise to an efficient and robust behavior descriptor. This method tested this algorithm in a number of domains against well established algorithms, and in all tests showed the best results.

3 EXISTING SYSTEM & PROBLEM DEFINITION

Various approaches have been proposed over the years for action recognition. On the basis of representation, they can be categorized as: time evolution of human silhouettes, space-time shapes, dense trajectories, and local 3D patch analysis, generally coupled with some machine learning techniques. All these works rely primarily on effective feature extraction. These feature extraction methods can be roughly divided into the following four categories: motion based, appearance based, space-time volume based, and space-time interest points or local features based. Motion based methods generally compute optical flow from a given action sequence, followed by appropriate feature extraction. However, these methods are known to be very susceptible to noise and easily lead to inaccuracies. Dynamic performance of a stand-alone wind and solar system with battery storage was analyzed. A few grid-connected systems consider the grid as just a back-up means to use when there is insufficient supply from renewable sources. To realize optimized applications in reconfigurable hardware, the high-level description has to be translated into a representation at the binary level.

- Self-Similarity Matrix
- Local 3D patch analysis
- Space-time interest point (STIP) based methods
- Multilinear independent component analysis (ICA) method
- Spatio-Temporal Features

3.1 Self-Similarity Matrix

SSM provides important insights in capturing the dynamics of a vector in both spatial and temporal dimensions, which is especially advantageous in a high dimensional space. We demonstrate the intuition of SSM by adopting the Lorenz attractor, which is a three-dimensional dynamical system representation that exhibits chaotic flow, noted for its figure-eight shape.

3.2 Local 3d Patch Analysis

We propose a new motion-adaptive sampling method of the silhouettes that favors fast moving body parts to better discriminate actions that only differ by small dynamic parts, like running and jumping. We introduce the 3D Distance Transform of Space Time Shapes to compute these fast moving parts. For each voxel in the Space-Time volume we compute the closest distance to the boundary. We adopt the 2D Distance Transform described in [6] and apply it to 3D.

3.3 Spatio-Temporal Features

Human behavior recognition, demonstrating the potential of a method based on spatio-temporal features in a domain where explicit shape models have traditionally been used. The behaviors their method can discriminate amongst, including walking, jogging, running, boxing, clapping and waving, are in fact well characterized by the reversal in the direction of motion of arms and legs. Hence these behaviors give rise to spatiotemporal corners, so the technique is well suited for dealing with their dataset.

4 MODIFIED SYSTEM & METHODOLOGY

Introduce some preliminaries on Self-Similarity Matrix. SSM provides important insights in capturing the dynamics of a vector in both spatial and temporal dimensions, which is especially advantageous in a high dimensional space. A SSM can be expressed by a N×N matrix Ri,j (//, v)= (//vi –vj//p),i, j ∈[1,N],where N is the length of a vector v, and // is a threshold distance. The threshold is a tuning parameter that changes the characteristics of the SSM dynamics. The // is a predefined norm, and the function (·) can be the Heaviside function (i.e. (x) =0ifx <0, and (x) =1 otherwise), or other proper filter functions .

4.1 Block Diagram

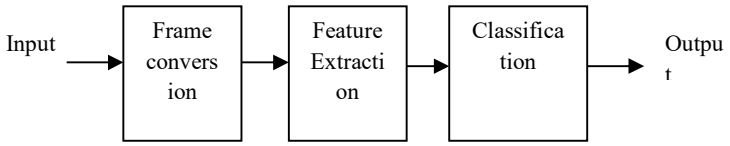


Figure 4.1: Block Diagram

Concretely, for a vector v =(v1,v2, ..., v_n), whose component vi is a column vector such that vi ∈ Rm×1andvi = (vi1,vi2, ..., vim)T, the SSM can be explicitly expressed

$$M_\epsilon(v) = \begin{pmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & 0 & d_{23} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \cdots & 0 \end{pmatrix},$$

Where dij is the distance between vector components vi and vj under distance norm p such that dij = vi −vj p. SSM behaves differently given different distance norm sp and the thresholds. In this project we set =0 for a complete representation for the j SSM representation which is defined below. We use M(v) to represent the resulting SSM of vector v though out this paper, and we use the p-norm defined by dij ={ mk=1|vik −vjk|p}1pas the distance metric. Note that the SSM holds the following three properties: Ri,j =Rj,i(symmetry); Ri,j ≤ 0 (non-negativity); and Ri,k <=Ri,j +Rj,k (triangle inequality).

Definition 1:The jSSM is defined as JRv,wi,j(v, w,v,w) = (v − vi −vj p1)(w − wi −wj p2), in which i, j ∈[1,N], v and w are two internal thresholds, p1and p2 are two distance norms. The jSSM will be used in our volume construction procedure. It defines an operation that generates one “fused” SSM out of two existing SSMs. The motivation for this extension is that, JRv,wi, jcan be viewed as defining the relationship between two trajectories, and represent their interaction in a uniform manner. We illustrate this intuition. The fusion of two three-dimensional trajectories (better be viewed in color). The first row shows the fusion between the trajectories generate by two Lorenz attractors, while the second row is for the fusion results between a Lorenz curve and a parameterized “Butterfly curve”. (1st column) the curve plots in 3D

space; (2nd column) the SSM for the blue trajectory;(3rd column) the SSM for the red trajectory; (4th column) the jSSM computed by SSM red ◦SSM blue; (5th column) the SSM computed by the gradient operator |SSMred−SSMblue|., in which the mutual dynamics between two different well-known trajectories are shown. We present two pairs of trajectories. The first pair is generated from two Lorenz attractors (1st row), and the second pair is generated from a Lorenz curve and a "Butterfly curve". For each pair, wecalculate its SSM, jSSM and "gradient SSM", which reveal subtle dissimilarity from different perspective. Now take the second pair of trajectories as an example. For the Lorenz curve L={lj}and the "Bufferfly curve"B={bi}, a recurrence will take place if a point lj on L returns to the neighborhood of a former point li, and simultaneously a point bj on the second trajectory B returns to the neighborhood of a former point bi. It is difficult to directly capture the mutual dynamics from trajectories L and B themselves, but by fusing two SSMs, the resulting SSM specifically encodes the mutual dynamics of the input trajectories. Meanwhile, it becomes convenient to make comparisons amongst different pairs

Algorithm 1 Joint-SSV Rank-1 Approximation

input : A 3-order tensor Joint-SSV $\mathcal{A} \in \mathbb{R}^{I\times J\times K}$, and an iteration termination threshold ϵ

output: Three vectors α, β, and γ that minimize $\|\mathcal{A} - \lambda\alpha\circ\beta\circ\gamma\|_2$, where $\alpha \in \mathbb{R}^{I\times 1}$, $\beta \in \mathbb{R}^{J\times 1}$, $\gamma \in \mathbb{R}^{K\times 1}$, and $\|\alpha\|_2 = \|\beta\|_2 = \|\gamma\|_2 = 1$

Initialize $\alpha^{(0)}, \beta^{(0)}, and\ \gamma^{(0)}$;

while $\|\mathcal{A} - \lambda^{(t)}\alpha^{(t)}\circ\beta^{(t)}\circ\gamma^{(t)}\|_2 \geq \epsilon$ **do**

$\tilde{\alpha}^{(t+1)} = \mathcal{A}\overline{\times}_2\beta^{(t)}\overline{\times}_3\gamma^{(t)}$;

$\tilde{\beta}^{(t+1)} = \mathcal{A}\overline{\times}_1\alpha^{(t)}\overline{\times}_3\gamma^{(t)}$;

$\tilde{\gamma}^{(t+1)} = \mathcal{A}\overline{\times}_1\alpha^{(t)}\overline{\times}_2\beta^{(t)}$;

$\alpha^{(t+1)} = \tilde{\alpha}^{(t+1)}/\|\tilde{\alpha}^{(t+1)}\|$;

$\beta^{(t+1)} = \tilde{\beta}^{(t+1)}/\|\tilde{\beta}^{(t+1)}\|$;

$\gamma^{(t+1)} = \tilde{\gamma}^{(t+1)}/\|\tilde{\gamma}^{(t+1)}\|$;

$\lambda^{(t+1)} = \mathcal{A}\overline{\times}_1\alpha^{(t+1)}\overline{\times}_2\beta^{(t+1)}\overline{\times}_3\gamma^{(t+1)}$;

end

The Weizmann dataset consists of videos of 10 different actions performed by 9 actors. Each video clip contains one subject performing a single action. The 10 different action categories are: walking, running, jumping, gallop sideways, bending, one-hand-waving, two-hands-waving, jumping in place, jumping jack, and skipping. Each of the clips lasts about 2 seconds at 25Hz with image frame size of 180×144.We evaluated two schemes, namely the JSSV-silh and theJSSV-hog3d, separately, using the two classification methods, respectively. The JSSV-pos scheme is unavailable for this dataset since there is no body joint point information available.

4.2 Advantages

- Generic and does not depend on the input feature vector
- Avoids the need to pre-align videos

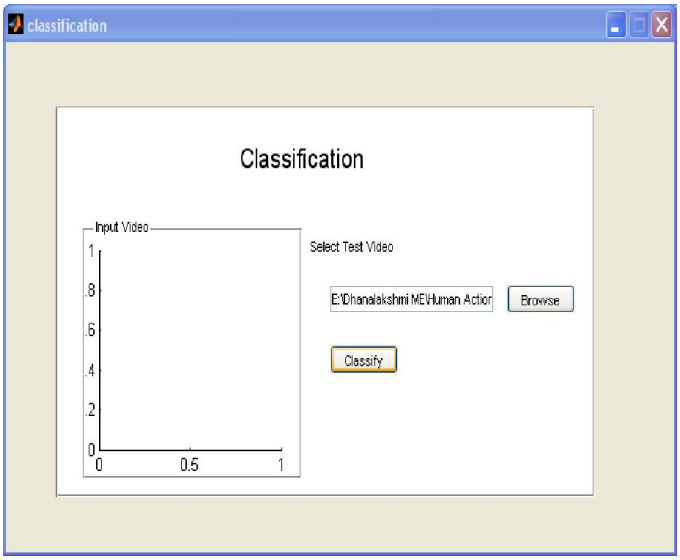
- Saving memory and computational time
- Effective method
- Simple system topology.

The calculation of the final decomposed vectors may take longer time for Als iteration to achieve the fitting error smaller than 10−10. The second component lies in the gaussian process based Classification for weizmann, kth, and ucf sports datasets, Because we scrutinized a large number of function combinations from a pool of function candidates (mean, covariance, Inference, likelihood). To some extent, the cost in the second Component is worthwhile, because it leads to removed class Bias, improved prediction robustness, and high recognition Rates. In our experiments, we performed all calculations using Ordinary lab-use computers. Actually, if parallel computing is Adopted, the computational burden in both the first and the Second components could be largely reduced.

5 RESULTS AND DISCUSSIONS

We evaluated our method on 5 well-known public datasets: Weizmann, KTH, UCF sports, CMU Mocap dataset, and the UCF-CIL dataset. Our goal is to evaluate the feasibility of our framework on various datasets with different Joint-SSV schemes. For all the results reported performed the recognition using the leave-one-out cross validation based on the two classification methods described. We perform Gaussian Process (GP) based classification using the following steps. First, for the 754 function combinations in the GP parameter settings, we perform our final action classification by randomly selecting 100 combinations. This is motivated by the fact that, in the absence of any prior information, all combinations are equally likely to provide the best classification rate. In addition, those 100 random combinations can largely cover almost all potential GP parameter combinations. Second, since each dataset has its own corresponding JSSV type, we perform action classification under these random 100 combinations for each type of JSSV ,and report our best rates for that type. Lastly, we make comparisons with both nearest neighbor classification and other methods reported in the literature.

Output For Walk



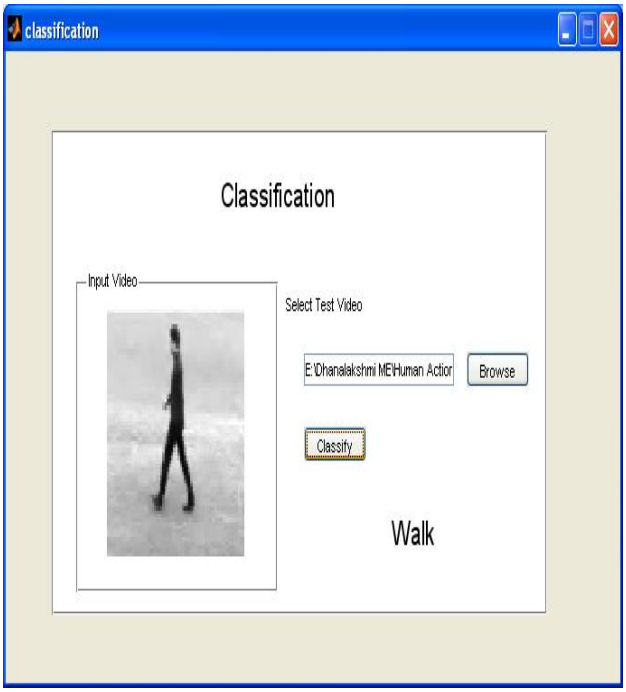


Figure 5.1: Output for classification of the action walk.

CONCLUSION

We use the recurrence plot theory to define a tensor representation of the dynamics of an action in video data, which we refer to as a Joint Self-Similarity Volume. We show that the Joint-SSV is sparse when applied to videos of human actions. In other words, it can be for the most part characterized by its rank-1 subspace representation. Therefore, by exploiting this sparseness, we reduce the high-dimensional recognition problem to a linear low-dimensional matching problem in a rank-1 subspace, without compromising our recognition accuracy. A particular feature of our approach is that it leads to a generic solution to this problem in the sense that our solution is independent of the type of input features, i.e. tracked points in a motion capture dataset, manually marked points, automatically extracted silhouettes, Histogram of Gradient (HoG) feature vectors, optical flow, etc. For reducing the dimensionality of the Joint-SSV, we introduced a new rank-1 tensor approximation algorithm that relies on an alternating least squares approach to find the optimal rank-1 decomposition. We demonstrate that in the case of Joint-SSV, the proposed decomposition largely preserve the salient characteristics of the scene dynamics. On the other hand, it leads to significant saving in both memory and computational time, since only a collection of rank-1 tensors is needed as the reference database for action class representation and matching. To validate our method, we devised three types of volume construction schemes, and performed experiments on five different public datasets. Also, since the proposed formulation is not dependent on the type of low-level features, the framework could be made view-independent by using view-invariant features. It is worth noting that the proposed new rank-1 tensor decomposition is general and may find an extensive set of applications beyond the action recognition problem explored in this paper, insofar as the data is of rank-1.

REFERENCES

1. Almustafa, K., Zantout, R., Obeid, H., Sibai, F., 2011. Recognizing characters in Saudi license plates using character boundaries. In: Internat Conf. on Innovations in Information Technology (IIT), pp. 415 –420.
2. Alvarez, S., Sotelo, M.Á., Ocaña, M., Llorca, D.F., Parra, I., Bergasa, L.M., 2010. Perception advances in outdoor vehicle detection for automatic cruise control.Robotica 28, 765–779.

3. Arnold, R., Miklos, P., 2010. Character recognition using neural networks. In: 11th Internat Symposium on Comput. Intell. and Informatics (CINTI), pp. 311–314.

4. Bailey, D., Irecki, D., Lim, B., Yang, L., 2002. Test bed for number plate recognition applications. In: Proc. of the First IEEE Internat. Workshop on Electronic Design, Test and Applications, pp. 501–503.
5. Brunelli, R., 2009. Template Matching in Computer Vision. Wiley. Caner, H., Gecim, H., Alkar, A., 2008. Efficient embedded neural-network-based license plate recognition system. IEEE Trans. Veh. Technol. 57, 2675–2683.
6. Dudarin, A., Kovaccic, Z., 2010. Alphanumerical character recognition based onmorphological analysis. In: IECON 2010 – 36th Annual Conf. on IEEE IndustrialElectronics Society, pp. 1058–1063.
7. Ganapathy, V., Lean, C., 2006. Optical character recognition program for images ofprinted text using a neural network. In: IEEE Internat. Conf. on IndustrialTechnology, ICIT 2006, pp. 1171–1176.
8. N. Benavides and P. Chapman, "Object oriented modeling of a multiple-input multiple-output flyback converter in dymola," IEEE Workshop on Computers in Power Electronics, pp.156-160, 2004.

9. Z. Ding, C. Yang, Z. Zhang, C. Wang, and S. Xie "A Novel Soft-Switching Multiport Bidirectional DC-DC Converter for Hybrid Energy Storage Systems," IEEE Trans. Power Electron., vol. 29, no. 4, pp. 1595-1609, April 2014.

10. S. Danyali, S.H. Hosseini, and G.B. Gharehpetian, "New Extendable Single-Stage Multi-input DC-DC/AC Boost Converter," IEEE Trans. Power Electron., vol. 29, no. 2, pp. 775-788, Feb. 2014.

11. H. Matsuo, W. Lin, F. Kurokawa, T. Shigemizu, and N. Watanabe, "Characteristics of the multi-input dc-dc converter," IEEE Trans. Industrial Electronics, vol.51, no.3, pp.625-631, June 2004.

12.Y.-M. Chen, Y.-C. Liu, and S.-H. Lin, "Double-input PWM DC/DC converter for high-/low-voltage sources," IEEE Trans. Industrial Electronics, vol.53, no.5, pp.1538-1545, Oct. 2006.