2D Information Space Based Action Recognition

L. Aneesh Euprazia, K.K. Thyagharajan

Abstract: Video based human action recognition has attained more attraction from the researchers and it predominates in the field of computer vision and pattern recognition. In this paper we deliver a new approach to suppress the background data and to extract 2D data of foreground human object of the video sequence. A combination of convex hull area, convex hull perimeter, solidity and eccentricity is used to represent the feature vector. Experiments are conducted on Weizmann video dataset to assess how the system is doing. The discriminative nature of the feature vectors assures accuracy in action recognition.

Keywords: action recognition, convex hull area, eccentricity, solidity

I. INTRODUCTION

Handling large video database is an important and challenging task and receives the attention of many researchers in computer vision and pattern recognition. The notable applications of human action recognition are human–computer interaction, surveillance, biometrics, virtual coaching, pedestrian tracking etc. Working with video datasets is challenging and good for the researchers so that their work can be extended for real time applications. Recognizing the actions in cluttered background, videos with brightness issues, different viewpoints and camera motion is interesting and challenging for the researchers. Representing a video as a two dimensional data like binary motion- energy image, motion history image helps to handle the video data easily. This representation scale down the dimensionality of the data resulting increase in efficiency. The 2D information space of a video is different for each action and the action is recognized by similarity measure and distance measure. While recognizing the action of a video, each frame may provide salient information. Researchers have used MHI(Motion History Image) to get 2D data from the video sequence. This is done by frame differencing the successive frames. So for each frame a difference operation is performed between current frame and previous frame. The frames are dependent on adjacent frames to get MHI. In this work the 2D data of video sequence is obtained by extracting the edge from each frame without depending on adjacent frames. Features such as area, perimeter, centroid, major axis length, minor axis length, compactness, solidity, eccentricity, roundness, convexity, elongation, bounding box, orientation

can be extracted from 2D data. Based on the type of application and nature of 2D data the features have to be opted. Extracting   features from static image data is extensively used in applications like leaf classification, fish classification, tumour classification etc. In this work a new combination of features is opted: convex hull area, convex hull perimeter, eccentricity and solidity. In the proposed method first step is to extract the 2D information of the video, second step is to derive the feature vectors from 2D data of the action videos and finally classify the actions. The key elements of the feature vector are convex hull area, convex hull perimeter, eccentricity and solidity. A neural network is used as the classifier to perform action recognition. The Weizmann video dataset is used to evaluate the performance of the proposed method. The forthcoming section of this paper is catalogued as follows: Section II provides literature survey, section III explains the proposed method, section IV talks about training and classification and section V concludes the paper.

II. LITERATURE SURVEY

Samy et al. [1] generates a sequence of difference images from successive frames of a video sequence. Features such as centroid, mean absolute deviation from the center of motion and intensity of motion are extracted from difference image resulting a set of features. Each video sequence is divided into overlapping time slices and the features obtained within each time slice is weighted and averaged to obtain a single attribute. The aggregated attributes of all the time slices of a video sequence is combined to obtain the final characteristic vector. The grouped features are handed to SVM for classification. Saumya et al. [2] uses set of frames as a block to generate Motion History Template and Binary Shape Templates (BSTs) for front, side and top view. MHT is obtained by subtracting consecutive frames in a block. BST is obtained by frame pixel-wise OR operation, for all the frames in the current as well as the previous blocks. HOG is produced for MHT and BST for three views and they are consolidated to get a final descriptor. SVM is used as the classifier.

Hany et al. [3] used CL1D, CL2D and CMHI submodels to handle 3D skeleton data, body part images and MHI respectively to recognize an action .CNN extracts the features from joint coordinates and body part images. 2D MHI is constructed by performing successive frame differences. The pixels are segregated as dynamic and static pixels by setting a threshold. Dynamic pixels in each frame is combined together to construct the 2D MHI. VGG16 pre-trained model extracts the features from 2D MHI. The output scores of sub-models are fused by Late Multiply Fusion

295

module and by obtaining the joint probability the action is recognized. Thanikachalam et al. [4] used MHI and Accumulated Motion Image (AMI). The outline information of AMI is extracted by Local Binary Pattern (LBP). Discrete Wavelet Transform (DWT) is used to bring out the spatial features from MHI. The central moments up to third order is calculated. The actions are classified by using KNN. Aaron et al [5] used binary motion- energy image [MEI] and MHI. Binarizing the MHI yields MEI.Hu moments are computed for all the images. By computing the mahalanobis distance of MEI the action is recognized. In Girum et al. [6] the shapes are represented as elements of the deformation space, which is a finite dimensional matrix Lie group. The shape retrieval system uses a distance metric. Shape analysis is done by k-means clustering. David et al. [7] trained a CNN to obtain 3D shape attributes of an image. Initially a 3DMap is estimated and then the Curvature, Contact and Volumetric Attributes are extracted from that map. A random forest classifier is used to discriminate the classes. Cem et al.[8] used contour information to extract the features like moment invariants, convexity, perimeter ratio, multiscale distance matrix, average margin distance and margin statistics. Linear Discriminant Classifier is used for classification.

Accordingly Motion history Image of a video is obtained by calculating the Frame difference for all successive frames in [1] [2] [3] [4] [5]. The drawback is the dependence of adjacent frames to get the MHI. As boundary outline extraction from a frame is independent of adjacent frames we can overcome the above issue. In [6], [7], [8] the work is complicated to attain 2D data. So by having simple and a new combination of features the work will become simple and action can be recognized accurately.

background for a given video input for each frame the intensity is set and then the edge information of the human object is extracted. The edges of all the frames are aggregated together to produce the aggregate edge. This obtained edge gives the 2-D information space of the video input. Fig.2 demonstrates the aggregate edge extraction process.
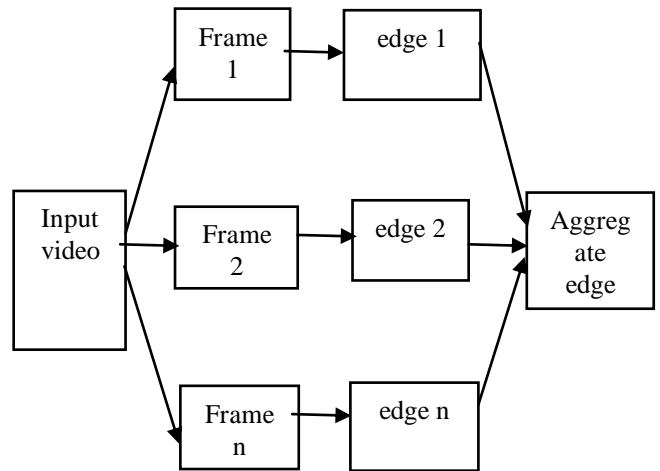

Fig.2. Aggregate edge extraction

The WEIZMANN dataset recorded ten actions (Bend, gallop side, jump, jump in place, jumping jack, one-hand wave, run, skip, two-hands wave, walk) performed by 9 subjects in a static background. It has 93 action video sequences.

Bend, skip, run and one hand wave actions are chosen for our research work. The Fig.1 shows the action frames of the video sequences of bend, skip, run and one hand wave.
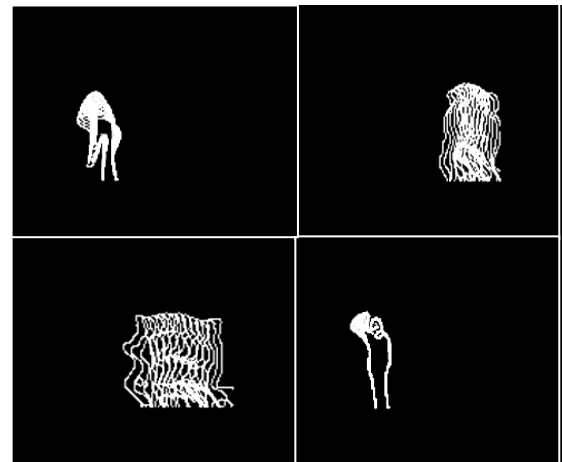

Fig.3. Aggregate edge of bend, skip, run and one hand wave actions of Weizmann dataset

Fig.1. bend, skip, run and one hand wave actions of Weizmann dataset

III. PROPOSED WORK

Handling large volume of data for processing prompts to high computational overhead. To overcome this issue, the video data is transformed into a two- dimensional information space. The spatial information plays a crucial role in action recognition. In this work the background is suppressed and the edge information of the foreground human object is extracted. In order to cut down the effect of

The video sequence of an action is given as input. The intensity value is set to 0.3 and then the edge of the foreground human object is extracted from all the frames. The set of obtained edges are aggregated together to generate the summed edge of the action video sequence. Fig.3 shows the aggregate edge of the actions bend, skip, run and one hand wave.

The derived attributes such as convex hull area, convex hull perimeter, solidity and eccentricity are used to determine the uniqueness of each actions.

A. Convex hull area

Let {b1,b2,…,bn} be the boundary pixels of the action shape. The convex hull is constructed by using the boundary points. The littlest convex polygon encompassing all the boundary points and connecting relevant boundary points is constructed. The convex hull area is obtained by the sum of the areas of each individual pixel within the outline of the convex hull of summed edge. Fig.4 shows the convex hull of the aggregate edges of the actions bend, skip, run and one hand wave.
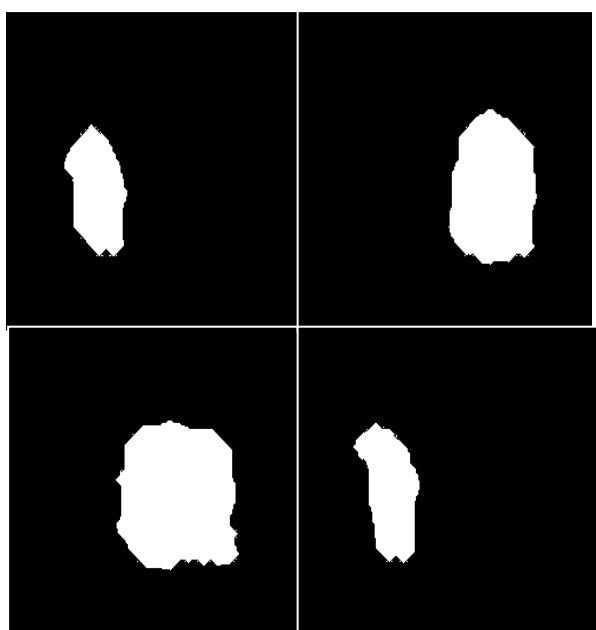

Fig.4. convex hull of bend, skip, run and one hand wave actions of Weizmann dataset

B. Convex hull perimeter

The convex hull perimeter is the total number of pixels on the contour of the convex hull. The pixel on the contour is categorized as horizontal, vertical and diagonal pixels. The horizontal and vertical pixels take unit distance, whereas diagonal pixel takes $\sqrt{2}$ distance. The total length is sum of number of horizontal pixels , number of vertical pixels and $\sqrt{2}$ * number of diagonal pixels.

$$\text{Convex hull perimeter} = \sum P_H + \sum P_V + \sqrt{2} * \sum P_D$$

C. Eccentricity

In an object : The ratio of the length of the minor axis to the length of the major axis is termed as eccentricity. The value of eccentricity has a standard bound and it lies between 0 and 1.

D. Solidity

Solidity measures the density of an object. A measure of solidity can be obtained as the ratio of the area of the summed boundary outlines to the area of a convex hull of summed boundary outlines of an action.

solidity = area / convex hull area

feature vector :

< convex hull area, eccentricity, solidity, length >

IV. TRAINING AND CLASSIFICATION

The feature values are measured in terms of pixels. The Table-I and Table II shows the feature values of the actions bend, skip, run and one hand wave performed by the actors 'Daria' and 'Denis' respectively of Weizmann dataset. The feature vector of each action possesses a different set of values which shows the discriminative property. When those vectors are fed in a classifier, the actions can be classified and recognized.

Table- I: Feature values of bend, skip, run and one hand wave actions of Weizmann dataset (person : Daria)

Action	Convex Hull Area	Eccentricity	Solidity	Convex hull perimeter
Bend	3878	0.889	0.9126	238
Skip	7468	0.786	0.9468	309
Run	10892	0.521	0.933	376
One hand wave	4023	0.930	0.879	263

Table- II: Feature values of bend, skip, run and one hand wave actions of Weizmann dataset (person : Denis)

Action	Convex Hull Area	Eccentricity	Solidity	Convex hull perimeter
Bend	3612	0.807	0.918	229
Skip	765	0.798	0.941	320
Run	10991	0.576	0.939	366
One hand wave	4029	0.939	0.857	271

First, from the video sequences the aggregate edge is extracted, then the features are extracted and finally features are concatenated together resulting feature vector. Second, a neural network is trained with the obtained properties. 70% of the feature vectors are used for training, 15% of the concatenated features are used for validation and 15% for testing. The training and testing process of the action recognition system is depicted in Fig.5. In the testing process

it is observed that the recognition rate of actions such as bend, run and one hand wave is good. Whereas the 'skip' action is misclassified as 'run' action for very few subjects. The recognition accuracy of the actions is given in Table III.

Table –III: Recognition rate of bend, skip, run and One hand wave actions of Weizmann dataset

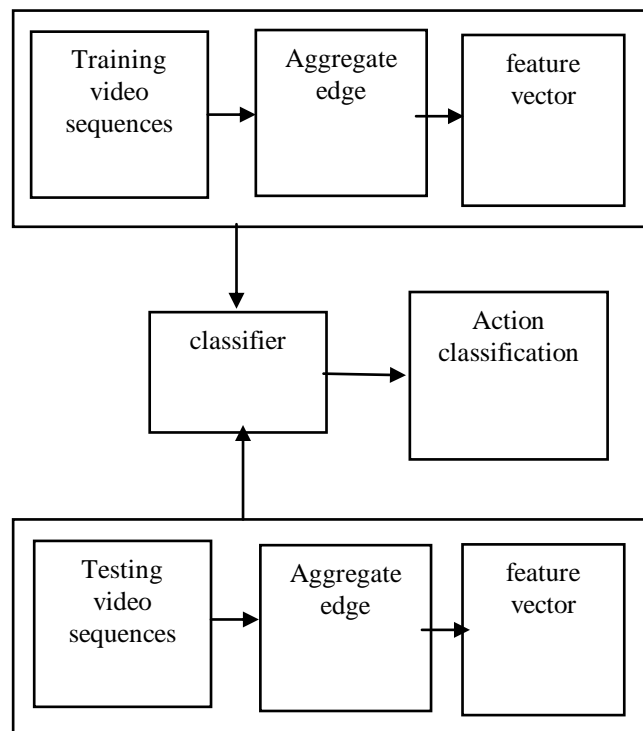| actions	bend	skip	run	One hand wave
accuracy	100%	89.1%	97.3%	100%

Fig.5. Training and classification of video sequences

V. CONCLUSION

In this paper, the feature extraction from video sequence is stated as (1) extracting 2D data, (2) extract convex hull area, convex hull perimeter, solidity and eccentricity, (3) concatenate features to generate feature vector. The values of the combined features generated from the videos of Weizmann dataset are discriminative in nature. Hence the system formulated by the proposed method is capable of recognizing the actions. The neural network trained with the obtained feature vectors recognize the action such as bend, run and one hand wave with good recognition rate.

REFERENCES

1. Samy Sadek, Ayoub Al-Hamadi, Bernd Michaelis & Usama Sayed , "A Fast Statistical Approach for Human Activity Recognition", International Journal of Intelligence Science , 2012,Vol.2 ,No.1
2. Saumya Jetley, Fabio Cuzzolin, " 3D Activity Recognition using Motion History and Binary Shape Templates"
3. Hany El-Ghaish , Mohamed E. Hussien, Amin Shoukry and Rikio Onai, "Human Action Recognition Based on Integrating Body Pose", Part Shape, and Motion, IEEE Access
4. V. Thanikachalam & K.K. Thyagharajan , "Human Action Recognition by Employing DWT and Texture" , Advances in Intelligent Systems and Computing 325, pp. 309-316
5. Aaron F. Bobick, James W. Davis, "Action recognition using temporal templates"
6. Girum Getachew Demisse , Djamila Aouada , and Bj€orn Ottersten , "Deformation Based Curved Shape Representation", 2017, Ieee Transactions On Pattern Analysis And Machine Intelligence, VOL. 40, NO. 6,
7. David F. Fouhey, Abhinav Gupta, Andrew Zisserman, "From Images to 3D Shape Attributes", IEEE Transactions on Pattern Analysis and Machine Intelligence,
8. Cem Kalyoncu , Önsen Toygar, "Geometric leaf classification", Computer Vision and Image Understanding, 2015,133 , 102–109

AUTHORS PROFILE

L.Aneesh Euprazia received the B.E. degree in Computer Science and Engineering from St.Xavier's Catholic College Of Engineering in 2009, received the M.E. degree in Software Engineering from Anna University, Trichy in 2011 . She is currently pursuing the Ph.D. degree in Information and Communication Engineering in RMD Engineering College under the supervision of Prof. K.K.Thyagharajan. Her research interests include video-based action recognition, computer vision, object segmentation, machine learning and deep convolutional neural network.

K.K. Thyagharajan received his B.Eng. degree in Electrical and Electronics Engineering from PSG College of Technology, Madras University, India and received his M.Eng. degree in Applied Electronics from Coimbatore Institute of Technology, India in 1988. He also possesses a Post Graduate Diploma in Computer Applications from Bharathiar University, India. He obtained his Ph.D. degree in Information and Communication Engineering from College of Engineering Guindy, Anna University, India in 2007. He is in teaching profession for more than three decades and served at various levels including Principal, Dean and Professor at various Engineering Colleges in Tamil Nadu-INDIA. During his tenure as a Principal and Dean he was a member of Governing Council of RMK Group of Institutions. He has written 5 books in Computing including "Flash MX 2004" published by McGraw Hill (India), which has served recommended as text and reference book by universities. He is a grant recipient of Tamil Nadu State Council for Science and Technology. He has been invited as chairperson and delivered special lectures in many National and International conferences and workshops. He is reviewer and editorial board member for many International Journals and Conferences. He is a recognized supervisor for Ph.D candidates and Master students at Anna University. He has published more than 100 papers in National & International Journals and Conferences. Nine candidates have completed PhD and eight more are doing PhD under his supervision. He has four patents published to his credit. His research interests include Computer Vision, Semantic Web, Image & Video Processing, Multimedia Streaming, Video Coding, Content-based Information Retrieval, Microcontrollers and e-learning. He is a Fellow of Institution of Engineers (India), life member of ISTE and CSI (INDIA), and also senior member and invited member in many professional associations. He has been recognized as a Teacher Par Excellence twice by the management of SSN institutions. He received Distinguished Faculty (Multimedia and Image Processing) Award from Venus International Foundation Chennai, Indo Global Education Excellence Award from International Benevolent Research Foundation Kolkata, and Best Administrator award from PEARL Foundation Madurai. He has been recognized by Marquis Who's Who in the World for his contribution to the technical society and his biography has been published in its 25th Anniversary Edition
E-mail: acdean@rmd.ac.in ,
kkthyagharajan@yahoo.com , kkthyagharajan@gmail.com
Scopus indexed publications:
https://www.scopus.com/authid/detail.uri?authorId=14046138000
Google scholar:
https://scholar.google.co.in/citations?user=R5jfVjoAAAAJ&hl=en
Research Gate: https://www.researchgate.net/profile/Thyagharajan_K_K
Publons (Web of Science):
https://publons.com/researcher/496086/thyagharajan-k-k