Recognition of Inter-Class Variation of Human Actions in Sports Video

Akila.K (✉ akilmoothy@yahoo.com)

R.M.K College of Engineering & Technology <https://orcid.org/0000-0003-0787-3130>

Research Article

Keywords: Human action recognition, Wavelet packet, video processing, Feature extraction.

Posted Date: July 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-684295/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Recognition of Inter-class variation of Human Actions in Sports Video

Akila.K

Assistant Professor, Dept. of CSE, R.M.K College of Engineering & Technology, Chennai, India
akilmoothy@yahoo.com

Abstract: Background: Human action recognition encompasses a scope for an automatic analysis of current events from video and has varied applications in multi-various fields. Recognizing and understanding of human actions from videos still remains a difficult downside as a result of the massive variations in human look, posture and body size inside identical category. **Objective:** This paper focuses on a specific issue related to inter-class variation in Human Action Recognition. **Approach:** To discriminate the human actions among the category, a novel approach which is based on wavelet packet transformation for feature extraction. As we are concentrating on classifying similar actions non-linearity among the features are analyzed and discriminated by Deterministic Normalized - Linear Discriminant Analysis (DN-LDA). However the major part of the recognition system relays on classification part and the dynamic feeds are classified by Hidden Markov Model at the final stage based on rule set.. **Conclusion:** Experiments results have shown that the proposed approach is discriminative for similar human action recognition and well adapted to the inter-class variation

Keywords: Human action recognition, Wavelet packet, video processing, Feature extraction.

1. Introduction

Human Action Recognition (HAR) and its applications are getting more attraction as the multimedia systems and its corresponding fields are getting more attention among the wide range of users. The field of HAR refers to the artificial intelligence in the computer system where a variety of algorithms is executed to recognize the actions [1]. Over the last two decade, researchers have explored application scenarios for HAR systems but not limited to, surveillance, healthcare, sports broadcast, machine/robotic control, machine-human / human-machine interaction, video retrieval and much other [1-4]. Apart from the potential application scenarios, the task of recognizing action is much more difficult due to varying camera viewpoint, lack of depth information, occlusion, overlapping objects, shadowed region and etc., [5]. The solutions to these problems are presented individually over the last decade, but still, the need for a promising approach is getting more attention in the research community. We target to the take problem of similar action recognition among the various sports activities. HAR system is emerging into the sports activities over the past few years [6]. The role of such systems finds its importance in broadcast systems and legal/illegal action monitoring. In general, most of the works are carried out either by orientation of joints or pairwise positions. The earlier approach is preferred when the sequence of actions are categorized as simple actions like standing in a queue, operating smart phones, speaking on a telephone and other similar actions carried out only with a part of the body. For complex actions (for instance, actions involved in sports) where the whole body acts in a faster motion, the latter approach is employed for higher accuracy [7-12]. Apart from these two approaches action localization has to be derived in an

effective way where the spatio-temporal locations of the actions are mapped. This stage of the HAR is most fundamental in major research works. Based on the source of input, HAR process is classified as frame-based approach and sequence-based approach. As the term defines, only the current frame of a video sequence is used to recognize the actions in the current frame. The influence of the reference image can be used or neglected depending upon the system process. For the sequence-based approach, both the current and initial frames are used to recognize the human action. The sequence based approach in HAR systems is more advantageous than the earlier one as it estimates the geometric displacement between the two processed frames for the action recognition [13]. Our proposed will also work based on the sequence-based HAR approach. Action recognition is a common problem among the researchers where the sequences of video frames are used for standard datasets. The problem here we address is about the similar actions among the sports activities, the problem is foreseen as a generic approach to classifying the legal and illegal action of play among the players. It is very much noted that when a player steps into a foul play, his action looks very similar to the legal action. For such instances, this proposed system uses Deterministic Normalized Linear Discriminant Analysis (DN-LDA) as an efficient action classifier to distinguish the similar actions in HAR systems. Further, this article is developed as follows. In section 2, related works in HAR systems is discussed in detail but not limited to similar action classification. Section 3 gives a detailed explanation on how the proposed is executed with DN-LDA. Section 4, briefs about the experimental setup and section 5 summarizes the results obtained along with conventional approaches. And finally, section 6 concludes the research work.

2. Related Works

Human action recognition is an important area of research in machine vision category [5]. Liu et.al, designed a framework for recognizing realistic actions from videos (Youtube videos) with motion cues to extract good features [14]. Also, this technique used Page rank technique to feature mining process. Which itself became a drawback that the mining of features get complex as the size of the data increases exponentially. In 2007 Marco et.al [15] presented a much similar work for security application combining audio recognition to the human actions in the surveillance systems. Analyzing the visual part, Gaussian model with time-adaptive per-pixel mixture is used to select the foreground (FG) scene. Auxiliary events in the foreground are detected by FG histogram model. This approach is able to achieve a classification accuracy level of 76%, which is just more than the average level of the techniques in the current trend. Towards solving the classification issues in recognizing similar human actions a better groundwork is presented [16] by Huan et.al. Among the two presented architectures, categorizing the framework is developed from a designer's point of view. To maintain the effectiveness of the system while handling large datasets, an adaptive feature selection algorithm can be appended to the architecture. This kind of additional feature will help the developer to upgrade their research work day-by-day as the technology evolves. When it comes to security based application in HAR systems, very vast set of research works were presented in the past decade [17, 18]. These techniques covered all sort of possible source of video input to deliver turnkey industrial solutions. In smart homes, human behavior is monitored for babysitters and old-age people monitoring. In these applications, the action recognition is limited only for unusual activities and there will be no much

confusion among the actions. The full objective of HAR system is not served in these kinds of environments. When the confusion level among the actions has increased the challenge for maintaining the accuracy of the similar actions increases. With the listed literature we develop our system for wide range of action recognition among similar events.

3. DN-LDA based HAR System

The process of recognizing the similar action in sports is carried out in four stages (i) pre-processing, (ii) feature extraction, (iii) feature selection and (iv) classification.

Fig 1. Basic function flow Diagram

3.1 Preprocessing

The very foremost stage in the HAR systems is the preprocessing stage. Initially, the frames from the input sequence are resized to 480x720 resolutions for smoother computations. In most of the conventional approaches the background and unwanted details are wiped out for faster processing. As we deal with similar actions in sports, we need qualifying information for the detailed classification. Working towards gathering detailed information, leaving the foreground section (where human and the objects are placed in the frame), and background section is scaled to gather detailed information. The floor where the action takes place will also help in classifying the sports with similar actions (field hockey vs. ice hockey; tennis vs. cricket). Another major work carried out in this stage is unifying the brightness level in all the frames. This is because; sports action will be recorded at different angles and different lighting conditions. Among all the methods histogram equalization is qualified as an effective method for answering the brightness issues.

Table 1. List of events grouped based on the similar actions involved

Group1	Group2	Group3	Group4
Tennis Cricket Baseball	Golf Skating Ice Hockey Field Hockey	Swimming Water Polo	Archery Shooting

3.2 Feature Extraction using Wavelet Packets

Once the frames are processed for unifying lighting conditions, occlusion and noise free frames, the next step of the HAR is to extract the valuable features for the classification. From the survey, it is explored that wavelet packets have higher efficiency in extracting features from grayscale images. The processed frame sequences are converted into gray scale format before executing the decomposition process. The mathematical model [wavelet reference] of Wavelet Transform is to decompose the temporal signal into a summation of the time-domain basis function of various frequencies.

$$\psi_{a,b}(t) = a^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad (1)$$

The above Eqn (1), is a decomposition of a function $f(t)$. This part contains a and b as scale and dilation parameter respectively. The special feature of Wavelet transformation is, it is totally reversible and the original signal can be reconstructed from its transformation form. Wavelet transform effectively splits a signal into low-frequency coefficients and high-frequency coefficients. The discrete form of wavelet coefficients is derived as

$$\langle s(t), \psi_{a,b}(t) \rangle = a^{-1/2} \int s(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (2)$$

Table 2. List of classified actions

Action Description	Action Code
Right Hand Shot above the waist	A01
Right Hand Shot above the knee	A02
Right Hand Shot to the ground level	A03
Left Hand Shot above the waist	A04
Left Hand Shot above the knee	A05
Left Hand Shot to the ground level	A06
Forehand return of serve	A07
Backhand return of serve	A08
Hockey - Hit	A09
Hockey - Push	A10
Golf – Tee Shot Swing	A11
Ice Hockey – Wrist Shot	A12
Skating	A13
Swimming Freestyle	A14
Swimming Butterfly	A15
Swimming Breakstroke	A16
Water Polo Throw	A17
Holding the Bow and Arrow	A18
Holding a Riffle	A19

On the other hand in wavelet packet transform, the decomposition is carried out in both scaling and dilating coefficients. This is performed in contrast to wavelet transform, where the decomposition is carried out only in scaling coefficients. Irrespective of the faster computation and detailed execution, the current trend is using 2-D wavelet packets. Here in our work, we consider 2-D wavelet packet decomposition. Considering the input functional space as then functional space is split into orthogonal space sub-bands as shown below in Eqn (3).

$$U = W_{0,0} \subset L^2(\mathbb{R}^2)$$

$$W_{i,[j_1,j_2]} = W_{i+1,[2j_1,2j_2]} \oplus W_{i+1,[2j_1,2j_2+1]} \oplus W_{i+1,[2j_1+1,2j_2+1]} \oplus W_{i+1,[2j_1+1,2j_2]} \quad (3)$$

In the above equation while replacing the initial vector values with $i=j$ as $0,1,2, \dots, n$ we get the below sequence

$$\left. \begin{aligned} W_{0,0} &= W_{1,[0,0]} \oplus W_{1,[0,1]} \oplus W_{1,[1,1]} \oplus W_{1,[1,0]} \\ W_{1,[1,1]} &= W_{2,[2,2]} \oplus W_{2,[2,3]} \oplus W_{2,[3,3]} \oplus W_{2,[3,2]} \\ W_{n,[n,n]} &= W_{n+1,[2n,2n]} \oplus W_{n+1,[2n,2n+1]} \oplus W_{n+1,[2n+1,2n+1]} \oplus W_{n+1,[2n+1,2n]} \end{aligned} \right\} \quad (4)$$

Also, the sub-bands engaged in the quad-tree splitting follow a separable product of 1D wavelet packet subspace. The next level of decomposing is to split the source image into eight quarter size images. As modeled in [19], wavelet sub band , 1, 2 is by the construction, the closure subspace spanned by the orthonormal system of wavelet packet functions:

$$\left. \begin{aligned} \tau_{8n}(a, b)^{q+1} &= \sum_k \sum_l h(k)h(l)\tau_{n,(k+2a, l+2b)}^q \\ \tau_{8n+1}(a, b)^{q+1} &= \sum_k \sum_l h(k)g(l)\tau_{n,(k+2a, l+2b)}^q \\ \tau_{8n+2}(a, b)^{q+1} &= \sum_k \sum_l g(k)g(l)\tau_{n,(k+2a, l+2b)}^q \\ \tau_{8n+3}(a, b)^{q+1} &= \sum_k \sum_l g(k)h(l)\tau_{n,(k+2a, l+2b)}^q \end{aligned} \right\} \quad (5)$$

The other four coefficients are the same replica of the first four coefficients as in Eqn (6). In order to achieve the best basis function representation, sub-bands are selected when the information cost is higher than the cumulative sum of consecutive next level sub-bands, instead of decomposing all of them.

$$\left. \begin{aligned} \tau_{8n}(a, b)^{q+1} &= \tau_{8n+4}(a, b)^{q+1} & \tau_{8n+1}(a, b)^{q+1} &= \tau_{8n+5}(a, b)^{q+1} \\ \tau_{8n+2}(a, b)^{q+1} &= \tau_{8n+6}(a, b)^{q+1} & \tau_{8n+3}(a, b)^{q+1} &= \tau_{8n+7}(a, b)^{q+1} \end{aligned} \right\} (6)$$

The advantage of the 2D wavelet packet transform to extract high range of qualifying features is achieved as a tradeoff with computational complexity. Further to reduce the computational complexity, recursive algorithm is used for deriving coefficient calculations as modeled in FFT [19]. In this approach, the computation complexity of the system is reduced to $\mathcal{O}(n/2)$.

Table 3. Flow of Processing

Vt - test video	I _b - luminous of background	F _r –Reference Frame	F _k - Key Frame
Step 1:	Preprocessing		
	Select F _k and F _r		
	Subtract F _k from F _r		
Step 2:	Illumination Enhancement		
	Compute illumination of Vt by I-I _b		
	Set Threshold value		
	Equal distribution of brightness level		
Step 3:	Feature Extraction		
	Use extractfeature(F _k ,D)		
	Define local and global descriptor		
	Apply Wavelet Transformation	$\langle s[t], \psi a, b(t) \rangle = a^{-1/2} \int s(t) \bar{\psi}(\frac{t-b}{a}) dt$	
Step 4:	Feature Selection		
	Define DR-LDA method		
	Estimate the highest eigenvalue		
	Decomposed to		
	$\left(\frac{1}{\lambda} s b - s w \right) = E D_{bw} E^T$		
	Where $E \in \mathbb{R}^{r_t \times r_r}$ - Eigenvector, $D_{bw} \in \mathbb{R}^{r_t \times r_t}$ -Diagonal matrix		
Step 5:	Classification		
	Use sports action rule sets		
	Apply HMM classifier $\lambda = \{A, B\}$		
	Where π - initial state		
	A - Transition probability matrix		
	B - Observation probability matrix		

3.3 Deterministic Normalized – Linear Discriminant Analysis for feature selection

The process of segregating related features among the subset of largely collected features from the input data is carried out in different approaches. Over the past decade [few references] many feature selection algorithms were developed based on computing and mining algorithms. Currently, the influence of machine learning and artificial intelligence, discriminant analysis like kernel discriminant analysis (KDA), generalized discriminant analysis (GDA) and linear discriminant analysis (LDA) are widely used. Recently in 2014 [24, 25] a new discriminant analysis approach named stepwise linear discriminant analysis (SW-LDA) for handling data that are non-linear in nature. Results shown in the article is promising for generic human action recognition system, but when it comes to solving the data with similar actions in the sports event, SW-LDA fails in achieving higher accuracy. Hence we propose, Deterministic Regularized – Linear Discriminant Analysis (DR-LDA) a new discriminant analysis for feature selection in HAR system. The major advantage of this is the use of the heuristic approach to estimate the regularization parameter. This regularization parameter is a small positive constant which is added to all the diagonal elements of within-class scatters matrix. Calculation the regularization parameter is initiated from Fisher's criterion $J^w(\alpha)$ in [26]. The mathematical model is solved to

estimate the highest eigenvalue of $\left(\frac{1}{\lambda} S_b - S_w\right)$, where S_b and S_w are within the class and between the class of scatter matrix respectively. Hence the Eigen Value Decomposition of $\left(\frac{1}{\lambda} S_b - S_w\right)$ is given by

$$\left(\frac{1}{\lambda} S_b - S_w\right) = E D_b w E^T$$

Here, $E \in \mathbb{R}^{r_t \times r_t}$ and $D_{bw} \in \mathbb{R}^{r_t \times r_t}$ are matrices of Eigenvector and diagonal matrix of corresponding eigenvalues respectively. In [27], DRLDA is evaluated for face recognition and cancer classification. Here, the performance of DRLDA to feature section in human action recognition is discussed in section 5.

3.4 Classifier

Hidden Markov model is used for learning the actions involved in the video sequences along with our classifier algorithm. Each stage of the HMM delivers a state transition symbol. In our approach actions are extracted into dynamic instants as meaning actions in section 3.2. Our HMM recognizes the actions in two different parts as training model and the probability of sequence recognition by the model. The HMM is characterized as $\lambda = \{\pi, A, B\}$. Here, π represents initial state distribution, A represents the matrix of transition probability and B represents a matrix of observation probability. We categorize the sport into the hidden state and the local information i.e., the action performed are the observation of the model.

3.5 Construction of Rule set

Group 1:

Basic Transition	: $q_0(P, O_p, O_F, AC) \xrightarrow{*} q_F(P_F, OP_F, OF_F, PA_F)$
Factors to be considered	: $\sigma(O_F, M_O, P_{pose}, E_{info})$
Constraints to be Analyzed	: Analyze {interaction (P, O_F), dynamic (M_O), Ppose (AC) }
Output	: Sports_Action label {K Φ (tennis_forehand, tennis_serve, batting, bowling, serve, strike)}

Group 2:

Basic Transition	: $q_0(P, O_p, O_F, AC) \xrightarrow{*} q_F(P_F, OP_F, OF_F, PA_F)$
Factors to be considered	: $\sigma(O_F, M_O, P_{pose}, E_{info})$
Constraints to be Analyzed	: Analyze {interaction((P, O_F)&(O_F, O_p)), static(M_O), Ppose(AC), type_Env(E_{info})}
Output	: Sports_Action label {K Φ (Hockey_hit, Hockey_push, skating, ice_hockey_shot)}

Group 3:

Basic Transition	: $q_0(P, O_p, O_F, AC) \xrightarrow{*} q_F(P_F, OP_F, OF_F, AC)$
Factors to be considered	: $\sigma(O_F, M_O, P_{pose}, E_{info})$
Constraints to be Analyzed	: Analyze {interaction(P, O_F), dynamic(M_O), Ppose(AC), type_Env(E_{info})}
Output	: Sports_Action label {K Φ (swim_freestyle, swim_butterfly, break_stroke, waterpolo_throw)}

Group 4:

Basic Transition	: $q_0(P, O_p, O_F, AC) \xrightarrow{*} q_F(P_F, OP_F, OF_F, AC)$
Factors to be considered	: $\sigma(O_F, M_O, P_{pose})$
Constraints to be Analyzed	: Analyze {interaction(P, O_F), obj_player_force(M_O), impact(O_F)}
Output	: Sports_Action label {K Φ (holding_rifle, holding_bow+arrow)}

q_0 – initial state q_F – Final state O_p –People Object O_F – Field object M_O —Object momentum

Pseudo code for Rule set sub-function:

<pre>interaction(P,O_F) { push_player(P,O_F) if(push_player(P,O_F)==1) return true; else return false; }</pre>	<pre>Ppose(AC) { find_pose(P,O_P,O_F) if (interaction(P,O_P) interaction(O_P,O_F) interaction(P,O_F)) return true with state of transition else return not in contact }</pre>
<pre>dynamic(M_O) or static(M_O) { V₀ be the disturbance caused object if (V₀ of O_F && (V₀ of O)_F==1) return state of current movement position of object else return its initial state (static) }</pre>	<pre>impact(O_F) { O_p produce another Object O₁ dynamic (M_O) of O₁ is found let F_O be the force applied if (newobj(F_O,M_O,O_P)==true) return raise_newobj else return no_newobj if(interaction(O₁,O_F) return target_fired & impacted else return target_notfired }</pre>

Fig 2. Pseudo code for Rule set

4. Experimental Setup and Simulation

In this section, we examine the accuracy and reliability of the proposed Human Action Recognition system for similar actions among various sports events. For the process of evaluation, we have collected various sports events available from YouTube and UCF sports dataset [31,32]. Considering all the sequences are of 20fps, we collect video sequence of length 3-5secs during which the action takes places. Simulation of the proposed system is carried out in 3.20 GHz Intel core i3 processor at Matlab 2013a environment. The limitation in the sequence length consider for the processing can be overcome, when the proposed approach is implemented in Supercomputers with huge data centers. Considering the data used in the experiment, sports events are selected from officially recognized games in Olympics and for adding more additional value one Non-Olympic sport is selected. The selected sports events are categorized into four different groups based on their similarity rate as shown in Table 1. For all the four groups, the proposed wavelet packet based DN-LDA is used for recognizing action among similar kind of sports events. Evaluation of the proposed scheme is discussed in the next section.

Currently, the proposed system is 11 different sports events which are grouped into 4 based on their similarity and more than 15 actions are framed for training and classification of the actions. This kind of classifications finds its application areas like broadcasting and monitoring the sports actions. In sports like cricket, tennis, and hockey, object tracking and recognition are used for judging the wickets, foul play or point count in the critical situation where human intervention is not able to quantify these situations [29].

Table 3. Confusion matrix for the actions in UCF sports and Youtube data

	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19
A01	98	1	0	0	0	0	1	0	-	-	-	-	-	-	-	-	-	-	-
A02	1	99	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
A03	0	0	100	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
A04	0	0	0	97	1	0	0	2	-	-	-	-	-	-	-	-	-	-	-
A05	0	0	0	1	98	0	0	1	-	-	-	-	-	-	-	-	-	-	-
A06	0	0	0	0	0	100	0	0	-	-	-	-	-	-	-	-	-	-	-
A07	0	0	0	0	0	0	100	0	-	-	-	-	-	-	-	-	-	-	-
A08	0	0	0	0	1	0	0	99	-	-	-	-	-	-	-	-	-	-	-
A09	-	-	-	-	-	-	-	-	100	0	0	0	0	-	-	-	-	-	-
A10	-	-	-	-	-	-	-	-	0	96	2	2	0	-	-	-	-	-	-
A11	-	-	-	-	-	-	-	-	1	3	96	0	0	-	-	-	-	-	-
A12	-	-	-	-	-	-	-	-	0	0	0	99	1	-	-	-	-	-	-
A13	-	-	-	-	-	-	-	-	0	0	0	1	99	-	-	-	-	-	-
A14	-	-	-	-	-	-	-	-	-	-	-	-	-	98	0	1	1	-	-
A15	-	-	-	-	-	-	-	-	-	-	-	-	-	0	98	2	0	-	-
A16	-	-	-	-	-	-	-	-	-	-	-	-	-	0	2	99	1	-	-
A17	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	99	-	-
A18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99	1
A19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	98

Similarly, HAR system with the ability to classify actions among similar actions, it can be used for automated online commentary / feeds portal and this can be used in find the actions of the player/athlete whether it is legal or illegal. For instance, in Group 3 there is four different types of events in swimming (i) freestyle (ii) butterfly (iii) backstroke and (iv) breaststroke. It is necessary to carefully watch the swimmers whether they are maintaining a style of the event. Apart from action recognition, here our system will be able to assist the organizers in judging the participants. For all the mentioned sports events, rule sets are framed with the basic knowledge about the sports [31]. When implemented along the sports organizers rule sets can be expanded to monitor and recognize the actions with more accuracy.

Fig 3. output snapshot for various group of sports action

5. Result and Discussion

The accuracy of the proposed system is estimated and compared with other conventional approaches and tabulated. This table only compares the proposed system with other approaches only in identifying the generic sports action. Addition to this, our system is capable of declaring detailed information on the particular action with is not available in other approaches [34-37]. The accuracy of the system is obtained by eqn.8.

$$Accuracy = \frac{TP}{\sum_{r=1}^p \sum_{c=1}^p C(r, c)} \quad (8)$$

the confusion matrix is further derived over several iterations in the experiments as actual action vs predicted actions in Table 3. The average accuracy rate of the system is compared with other HAR systems and it is well justified that our system outperforms all the other conventional systems.

Table 4. Recognition accuracy of HAR systems comparison

Fig 4. Recognition Accuracy

Method	Accuracy (%)
Dense Trajectories [34]	85.5
Multi-ch. Gabor + HOG3D [36]	85.6
Multi-ch. Gabor + SOD [37]	87.5
HBRT [35]	90.8
HBRT/VOC [35]	96.2
Our Method	97.4

6. Conclusion

The human action recognition method discussed in this paper is novel in classifying similar actions in sports events. The importance of generic human action recognition has its own applications is security and machine learning. Our research work opens a platform for broadcasting application and monitoring athlete's legal/illegal actions in the sports events. With a trained dataset and rules framed with the end user, our intelligent HAR system is capable of achieving the accuracy rate of 97.4% which is higher than the other state of art approaches. So far in the experiments and literature it is inferred that DN-LDA is the most promising technique for feature selection process. As future advancements, the accuracy of the system can be increased by adopting different feature extraction and classification approach.

Data Availability:

The sports video data used in this study are from benchmark dataset publicly available in online at the repository https://www.crcv.ucf.edu/data/UCF_Sports_Action.php

Declarations:**Funding:**

Not received any funding from the external source.

Conflicts of interest/Competing interests:

The Author declares that there is no conflict of interest'.

Availability of data and material:

No data published anywhere; test videos are collected from the standard repositories.

Code availability:

Implemented using matlab, not deposited the code in online repositories

Authors' contributions:

Single person fully responsible to the work content.

References

1. J. Aggarwal and M. Ryoo, (2011). Human activity analysis: A review, *ACM Comput. Surveys*, 43(3), 16-1:16-43.
2. Ali S, Shah M (2010). Human action recognition in videos using kinematic features and multiple instance learning, *IEEE Trans Pattern Anal Mach Intell (TPAMI)*, 32(2), 288–303.
3. Bobick AF, Davis JW (2001). The recognition of human movement using temporal templates, *IEEE Trans Pattern Anal Mach Intell (TPAMI)*, 23(3), 257–267.
4. Gorelick, L. Blank, M. Shechtman, E. Irani, M. Basri, R, (2007). Actions as Space-Time Shapes, in *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 29(12), 2247-2253.
5. J. M. Chaquet, E. J. Carmona, A. Fernandez-Caballero, (2013). A survey of video datasets for human action and activity recognition, *Computer Vision and Image Understanding*, 117(6), 633 – 659.
6. S. Sadanand, J. J. Corso, (2012). Action bank: A high-level representation of activity in video, in: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1234–1241.
7. Yin, J, Yang, Q, and Pan, J. J, (2008). Sensor-based abnormal human activity detection. *IEEE Trans. Knowl. Data Eng*, 20, 1082–1090.
8. C. Zhang, Y. Tian, (2012). RGB-D camera-based daily living activity recognition, *J. Comput. Vis. Image Process*, 2(4), 12.
9. Kajita, H. Hirukawa, K. Harada, K. Yokoi, (2014). *Introduction to Humanoid Robotics*, Springer, Berlin Heidelberg.

11. W. Lin, H. Chu, J. Wu, B. Sheng, Z. Chen, (2013). A heat-map-based algorithm for recognizing group activities in videos, *IEEE Trans. Circuits Syst. Video Technol*, 23(11), 1980–1992.
12. T.L.M. van Kasteren, G. Englebienne, B.J.A. Kröse, (2011). Human Activity Recognition from Wireless Sensor Network Data: Benchmark and Software, *Activity Recognition in Pervasive Intelligent Environments*, Atlantis Press, 165–186.
13. F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, R. Bajcsy, (2014). Sequence of the most informative joints (SMIJ): a new representation for human skeletal action recognition, *J. Vis. Commun. Image Represent*, 25(1), 24–38.
14. Xia, Li-ming, Jin-xia Huang, and Lun-zheng Tan, (2013). Human action recognition based on chaotic invariants, *Journal of Central South University*, 20, 3171-3179.
15. J. Liu, J. Luo and M. Shah, (2009). Recognizing realistic actions from videos ‘in the wild’, *Proc. IEEE CVPR*, 1996 -2003.
15. M. Cristani, M. Bicego and V. Murino, (2007). Audio-visual event recognition in surveillance video sequences, *IEEE Trans. Multimedia*, 9(2), 257 -267.
16. H. Liu and L. Yu, (2005). Toward integrating feature selection algorithms for classification and clustering, *IEEE Trans. Knowl. Data Eng*, 17(4), 491 -502.
17. O. Brdiczka, M. Langet, J. Maisonnasse and J. L. Crowley, (2009). Detecting human behavior models from multimodal observation in a smart home, *IEEE Trans. Autom. Sci. Eng*, 6(4), 588 -597.
18. K. K. Reddy and M. Shah, (2012). Recognizing 50 human action categories of web videos, *Mach. Vis.Appl*, 24(5), 971 -981.
19. M. Misiti, Y. Misiti, G. Oppenheim, (2007). *Wavelet Toolbox 4 for Use with Matlab*, The math works Inc.
20. L. Wang and D. Suter, (2007). Recognizing human activities from silhouettes: Motion subspace and factorial discriminative graphical model, *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 1 -8
21. Wang, Bin, (2013). Multi-scale locality-constrained spatiotemporal coding for local feature based human action recognition, *The Scientific World Journal* 2013.
22. Cai, Jia-xin, Xin Tang, and Guocan Feng, (2014). Learning pose dictionary for human action recognition, 2014 22nd International Conference on Pattern Recognition (ICPR). IEEE.
23. Gu, Feng, (2014). *A Multiple Kernel Learning Based Fusion Framework for Real-Time Multi-View Action Recognition, Ambient Assisted Living and Daily Activities*. Springer International Publishing, 26-33.
24. Siddiqi, Muhammad Hameed, (2014). Video-based human activity recognition using multilevel wavelet decomposition and stepwise linear discriminant analysis. *Sensors* 14(4), 6370-6392.
25. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, (2014). Large-scale video classification with convolutional neural networks, in *IEEE Conf. Computer Vision and Pattern Recognition*, 1725-1732.
26. Mika, S. Ratsch, G. Weston, J. Scholkopf, B. Mullers, K. (1999). Fisher discriminant analysis with kernels, *Proceedings of the 1999 IEEE Signal Processing Society Workshop Neural Networks for Signal Processing IX*, Madison, WI, USA, 23–25, 41–48.

27. A. Sharma and K. K. Paliwal, (2015). A deterministic approach to regularized linear discriminant analysis, *Neuro computing*, 151(1), 207–214.
28. Y. Du, W. Wang and L. Wang, (2015). Hierarchical recurrent neural network for skeleton based action recognition, in: *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1110–1118.
29. P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, (2010). Object detection with discriminatively trained part-based models, *IEEE Trans. Pattern Anal. Mach. Intell.* 32(9), 1627–1645.
30. Ronchetti, Franco, (2015). Distribution of action movements (DAM): a descriptor for human action recognition, *Frontiers of Computer Science*, 9(6), 956-965.
31. Mikel D. Rodriguez, Javed Ahmed, and Mubarak Shah, (2008). Action MACH: A Spatio-temporal Maximum Average Correlation Height Filter for Action Recognition, *Computer Vision and Pattern Recognition*.
32. Khurram Soomro and Amir R. Zamir, (2014). *Action Recognition in Realistic Sports Videos, Computer Vision in Sports*. Springer International Publishing.
33. Shi, Qinfeng, (2011). Human action segmentation and recognition using discriminative semi-markov models, *International journal of computer vision*, 93(1), 22-32.
34. Wang, H, Klaser, A, Schmid, C, Liu, C, (2013). Dense trajectories and motion boundary descriptors for action recognition , *Int. J. Comput. Vis.*, 103(1), 60 -79.
35. Al Harbi, Nouf, and Yoshihiko Gotoh, (2015). A unified spatio-temporal human body region tracking approach to action recognition, *Neurocomputing*, 161, 56-64.
36. I. Everts, J. C. van Gemert, and T. Gevers. (2013). Evaluation of color STIPs for human action recognition, In *CVPR*.
37. H.Zhang, W.Zhou, C.Reardon, L.E.Parker, (2014). Simplex-based 3d spatio-temporal feature description for action recognition, in :*Computer Vision and Pattern Recognition (CVPR)*, 2014 IEEE Conference on, 2067–2074.