Jump Classification with Age and Gender Detection

Premanand Ghadekar [1*], *Riddhi* Karandikar[1], *Janhvai* Pawar[1], *Prithviraj* Sangle[1], *Parth* Patil[1] and *Parul* Patle[1]

[1]Department of Information Technology Vishwakarma Institute of Technology Pune, India

Abstract. The major issue is being able to identify human behaviour. The main issue for video categorization systems is common human actions in videos. For instance, a running motion will be included in a long jump or running sports film. Due to its multiple applications in areas like person monitoring, human-to-object interaction, and more, human action recognition is a crucial study subject in the science of computer vision. The computer vision community finds the video classification problem to be very difficult. The main reason that the video categorization problem is so challenging is the shared activities that are seen in the video. A high jump sports film, for instance, combines two distinct actions—running and high jumping—that are also shown in other videos, like running or hurdling sports videos. With just one frame that captures the specific action of the event, the human brain can quickly identify the correct occurrence in a film. By removing a few significant frames from the video and using those frames to conduct the classification procedure, the same premise may also be used in video classification systems.

1 Introduction

Human activity detection is a technique for anticipating someone's behaviours based on a specified order of that person's actions and outside variables. A variety of social applications, such as intelligent surveillance, intelligent robotics, and other tracking systems, are impacted by the difficulty of activity detection. Human activity recognition is a technique for analysing human movements using computer and machine vision technology (HAR). Sensors can capture motion that is seen as coming from humans, such as actions, gestures, or behaviours. The movement data is then translated into human activity recognition code, which computers can understand and carry out, to create the action order. It is a specific type of time series classification problem that requires correctly classifying the action that is being performed. Human activity recognition, or simply "HAR," is a wide field of study that focuses on identifying a person's precise movement or action using sensor data. Common indoor

* Corresponding author: premanand.ghadekar@vit.edu

activities include walking, talking, sitting, and standing Hare included in movements.

For numerous uses, including item detection, environmental change, and agriculture, land use/land planning, urban planning, surveillance, geographic mapping, and disaster management, image categorization in remote sensing pictures is crucial. The process of classifying and labelling groups of pixels or vectors within an image following specified guidelines is known as image classification. Both supervised and unsupervised categorization is possible. The process of categorizing a particular object in an image is referred to as image classification. This method's primary goal is the accurate detection of visual features. Both supervised and unsupervised methods are used to categorize images. This approach provides supervised classification results based on the decision boundary generated, which mostly depends on the input and output supplied during model training. However, the unsupervised classification does not directly feed features into the models; instead, it analyzes the input dataset to provide results.

Choosing an acceptable classification system, feature extraction, selecting quality training samples, choosing an ideal classification method, post-classification processing, and ultimately evaluating overall accuracy are the primary phases in image classification systems. The most prominent neural network model that we can use to solve the picture classification problem is convolutional neural networks (CNNs). The location of items in a picture, also known as object localization, and the categories to which they belong. For the classification of gender, The CNN approach was put to use, as it can be used for facial detection. The HAAR Cascade classifier is used for face detection. This classifier, localizes the facial area for detection, while the CNN has 2 parameters in the output layer, Male and Female which are used to classify the gender of the person.

The next-generation image and video processing systems have successfully adopted this technology, which may be used to look for a certain class of objects like automobiles, people, animals, etc. Only with the introduction of deep learning approaches have the most recent improvements in this methodology been made available. Therefore, for the high jump, long jump, and pole vault classification, we will use videos.

2 Literature Review

The fundamental challenge in this paper's investigation of human behaviour recognition is bridging the semantic gap between the symbolic domain of human interpretation and analogy observations of the physical world. Human motion analysis is a prominent area of study in the communities of computer vision and video processing. The article suggests a brand-new architecture for identifying online human activity in sporting event recordings. Since the suggested method uses the TBM framework rather than conventional probability theory, it is unique. Belief functions, which are more adaptable than probabilities, are used to build this framework. [1]. It primarily concentrates on modelling the human body, high-level recognition algorithms with domain expertise, and the amount of detail required to comprehend human activities. Knowing what is happening in the situation is the ultimate goal of learning human motion. High-level scene and context knowledge is part of the objective [2]. Computer analysis of human behaviour is becoming a subject of increasing interest. The act of capturing the motion, also known as human motion capture, is a crucial component of this endeavour. Applications where one or more subjects are tracked over time and maybe watched for specific actions fall under the surveillance category. The surveillance of a parking lot is a well-known example, where a system monitors individuals to determine whether they may be going to conduct a crime, such as car theft [3]. Video human action recognition is a difficult topic with many potential solutions. In this study, a straightforward representation is suggested that is intended to reflect such motion interactions. To describe motion information and make the final representation resistant to camera movement, we use both global and local reference points. In computer vision, there is currently considerable study on the problem of human activity recognition in videos. Years have seen significant advancements, particularly with the development of the bag-of-features architecture and local invariant features. We have outlined a method for motion-based action modelling in this study [4] [5]. The important frames from a video are extracted, classified using CNN, and the predictions for all the extracted frames are averaged in this study's method for video recognition. The concept behind the system was drawn from how people categorize events in videos, where the most significant or captivating frames have a greater influence on the choice of event category than other, less consequential frames [6]. Applications for human motion analysis can be found in a wide range of fields, including sports analysis, surveillance, and content-based image storage and retrieval. Detecting, locating, and identifying humans as well as recognizing human activity are the key scientific difficulties in human motion analysis. In this study, they covered topics such as pole identification, motion analysis of human shape, and human shape. The Transferable Belief Model (TBM)-based framework for human motion analysis has been shown to perform well in identifying athletes' motions and activities. The TBM enables the representation of uncertainty and conflict that is not possible in conventional probability theory. The Belief State Scheduler, which in this study recognized actions as a series of understandable acts, and the Temporal Credal Filter, which smooths belief on actions, both completely use these concepts. [7][8]. Finding groupings of athletic disciplines that have an impact on men's decathlon performance is the issue. The clustering techniques for categorizing groupings of sports disciplines in the decathlon structure were employed in this paper. Using hierarchical models, groups of athletic disciplines that influence sports performance can be identified [9] [10]. In this paper, we provide a tool termed the temporal credal filter with conflict-based model modification for online smoothing of belief functions in the transferable belief model (TBM) framework. To combine beliefs, the TCF-CMC explicitly models conflict information in TBM and considers temporal characteristics of belief functions [11].

This study demonstrates the value of using temporal structures to distinguish between simple and sophisticated human actions. Other sorts of contextual data as well as richer video representations are other future directions. The detection of the launch frame in long jump videos is the issue. MATLAB is the program utilized, and the OpenPose method is what builds the project [12]. The issue is centred on the use of computer vision in sports analysis, ball tracking, and player monitoring. For this project, no hardware is used at all. The benefit is that it expands the scope of applications for resumes. Convolutional and recurrent neural networks are used to categorize 15 different sports, which is the issue. The program is called Vs-code. CNN, RNN, GRU, and VGG-16 are the algorithms that are employed [13] [14]. Object detection and identification is the issue. The vs-code program is also utilized here. Long Jump Mathematical Analysis is the issue. No software is utilized here. Using computer vision to detect and track things is problematic. Single Shot Detector (SSD) and Mobile Nets algorithms are the ones that are utilized. The benefit is an average real-time detection level for all items above 99%. The issue is Python-based face detection, eye detection, and Haar features. The Cascade Classification algorithm is employed. Finding objects in videos to extract is the key issue. Background separation and Haar Cascade methods are employed in the algorithm. Long Range Detection offers an advantage [15].

A group of related computer vision tasks, including finding things in digital photos, are referred to as object recognition. Predicting the class of one object in an image is one example of the actions involved in image classification. To draw a bounding box around the extent of one or more objects in an image, their location must first be determined. This process is known as object localization. Combining these two objectives, object detection locates and categorizes one or more things in an image [16].

3 Methodology

This section discusses the entire process of the development of our jump classification, gender, and age detection. The selection of the dataset, data pre-processing, development, and training of the model, along with the evaluation of its performance, are all done using the conclusions gained from the study of related literature.

3.1 Data Collection

No direct form of the dataset was available. We collected the data set from various sources on the internet like YouTube. There are 4 types of jumps in our project, namely the High jump, triple jump, pole vault, and long jump. We downloaded about 300 videos in total for all 4 types of jumps. All these videos were in .mp4 format, from YouTube. The lengths of all the videos are not more than 2 minutes each video. This was the dataset used for classifying which type of jump it was out of the 4.

For the identification of the gender and classification whether the person is a male or a female; we used an image dataset. This dataset was extracted from all the downloaded videos itself. We used the library available in python to extract images from the video. A new file was created that stored all these images which were later on used for classification.

3.2 Data Distribution

The total dataset of 6000 images is compiled for object classification and detection which were captured from cameras and some of them were collected from the internet. Out of all images, 3600 images were used for training and 2400 images were used for testing.

Table 1. Distribution of Dataset

Sr. No.	Jump Type	Training	Testing
1	High Jump	65	25
2	Long Jump	63	20
3	Pole Vault Jump	57	18
4	Triple Jump	35	17
5	Total Number of Videos: 300		

3.3 Data Pre-processing

The dataset for each of the four was first analysed. A list was created where the videos of all 4 types of jumps were stored. The data was split into training and testing data. The data set was converted into a .csv file. The videos were later fed to the network and stored in an array form, features were extracted from the video and labelled. The images, captured from the video were stored in a separate folder labelled as 30_0, i.e. Age, gender. 0 is for males and 1 is for females.

3.4 Proposed Methodology

3.4.1 Algorithm

Input – Videos of different types of jump

Output – Classifies the videos into different types

1. Import required libraries
2. Prepare the training and testing data
3. Convert the data to .csv file
4. Extract the frames and resize them
5. Feature Extraction
6. Label Encoding – label the data
   1 → High jump
   2 → Long jump
   3 → Pole vault
   4 → Triple jump
7. Define Hyper parameters –
   IMG_SIZE = 224
   BATCH_SIZE = 64
   EPOCHS = 100
   MAX_SEQ_LENGTH = 20
   NUM_FEATURES = 2048
8. Feed this data to a sequence model (CNN)
9. Inference

3.4.2 Flow Diagram

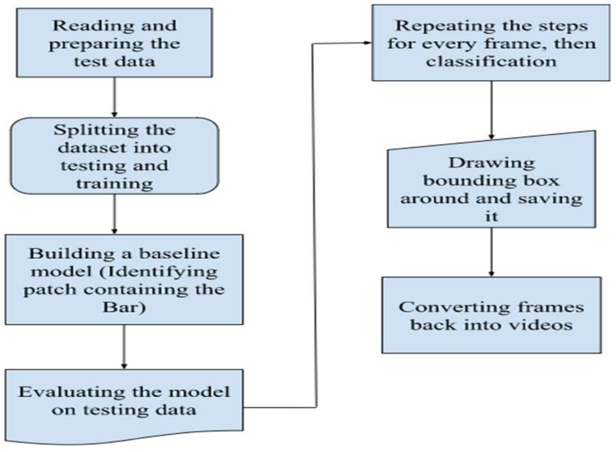

Fig. 1. Project flow diagram for jump classification with age and gender detection model.

https://doi.org/10.1051/itmconf/20235302003

3.4.3 Proposed System Diagram

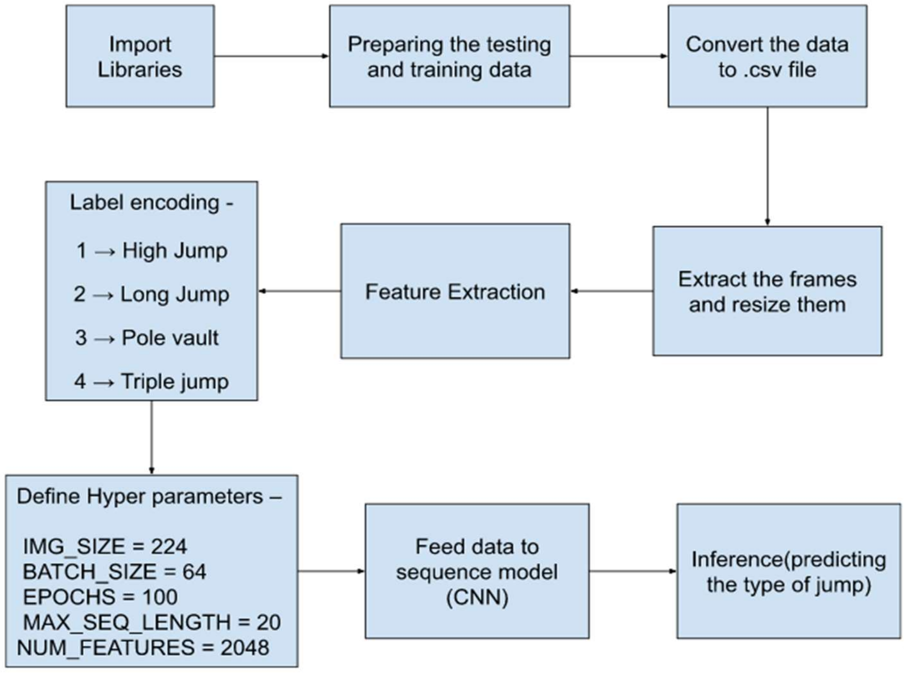

Fig. 2. The proposed model for Jump Classification with Age and Gender Detection.

4 System Description

This paper intends to classify the four different types of jumps, High jump, long jump, Pole Vault jump, and Triple Jump. Various python modules have been used to train the model or create the dataset, and extract features from the video for creating the model. Sample videos are installed in the training dataset which is referred to or compared with the test videos given to the system. Since jumping is an action then a comparison between different videos shall be done. Image comparison becomes easy as it has two single frames. In this paper, when it comes to video comparison the videos are accessed with python function video capture so multiple frames or images are captured of the video and since each image may give a different output than a collective output is obtained using the concept of CNN. Also, there can be different numbers of frames in different videos so it becomes difficult to compare frames. Here the concept of padding is used. (Padding is the concept where null or empty frames are added in the video to make the number of frames in both videos equal.) Comparison is done based on the features of the image. The features of each frame are different and are extracted using the keras function of the Tensor Flow module. A comparison of these images is done based on these features.

For the detection of the age and gender of the person in the video, frames are extracted using object detection, which is used later on for prediction purposes.

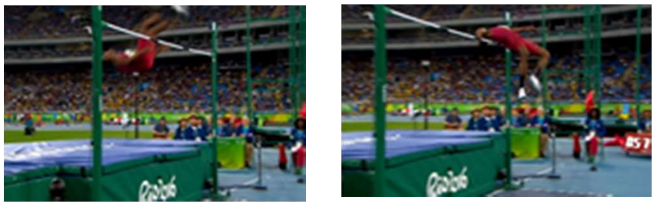

Fig.3. High Jumps

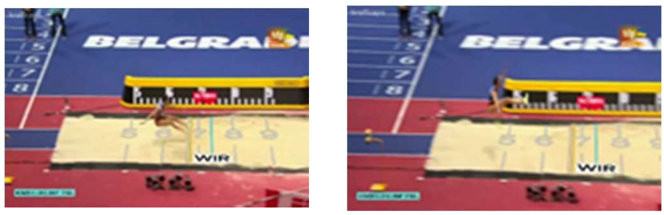

Fig.4. Long Jumps

Fig.5. Pole Vault Jumps

Fig.6. Triple Jumps

5 Results

The Adam optimizer is used to build the model once it has been developed for the classification of the different sorts of jumps. It is a highly effective optimization approach used for deep neural network training. The percentage of the type of jump in the video is 20.13% for High Jump, 19.97% for Long Jump, 19.97% for Triple Jump, and 19.97% for the pole vault. The overall accuracy for the model is 85.93%. After the type of jump has been classified, the gender and the age of the person are also detected with an accuracy of 93.55%. Our model was successfully implemented with all the objectives satisfied.

Table 2. Accuracy and outcome of the type of jumps.

Type of jumps	Dataset size	Accuracy	Prediction
High and Long Jump.	169	92.84%	High Jump: 60.24% Long Jump: 39.76%
High Long and Triple Jump.	298	89.56%	High Jump:42.68 Long Jump:29.94 Triple Jump: 27.38
High Long Triple Jumps and Pole Vault	482	85.93%	High Jump: 20.10% Long Jump: 20.07% Pole Vault: 19.94% Triple Jump: 19.94% [UNK]: 19.94%

Table 3. Model Accuracy for jump classification and age and gender detection.

Task Performed	Accuracy	Precision	Recall	F1 score
Classification of the jump model	85.93%	89.03%	89%	89.03%
Age and Gender	93.55%	94.30%	94.30%	93%

6 Conclusion

Since these jumps are quite similar, sometimes errors May appear because the body positions and postures of the jumps are quite similar. If the long jump (Fig 4) and triple jump (Fig 6) are considered then the most common similarity in the frame is sand. If the frames are not divided properly during the time of hop then the results will not be good. The same happens with the vault and high jump since the body postures are similar. If the pole is properly detected then the results will be accurate. Otherwise, the results might be confusing between the high and vault jump. Along with classifying what kind of jump it is out of the 4 mentioned jumps, the model also detects the age and the gender of the person in the video. The detection of age and gender in the model as mentioned before is done by using various libraries in python.

References

1. Drücker, Sven, Konrad Schneider, Navneet-Kaur Ghothra, and Swantje Bargmann. "Finite element simulation of pole vaulting." Sports Engineering volume 21, issue 85-93, 2017
2. Ramasso, E., Rombaut, M., & Pellerin, " State Filtering and Change Detection Using TBM Conflict Application to Human Action Recognition in Athletics Videos", IEEE Transactions on Circuits and Systems for Video Technology vol 17 issue 7, 2007
3. Arampatzis, Adamantios, Falk Schade, and Gert-Peter Brüggemann. "Effect of the pole–human body interaction on pole vaulting performance." Journal of biomechanics vol 37, issue. 9, 2004
4. Tözeren, Aydın. "Energy Transfers: In Pole Vaulting, Running, and Abdominal Workout.", Human Body Dynamics: Classical Mechanics and Human Movement, vol 12 issue 2, 2005.
5. Fukushima, Toshihiko, Satoshi Nishikawa, and Yasuo Kuniyoshi. "Active bending motion of pole vault robot to improve reachable height." IEEE international conference on robotics and automation (ICRA)vol 4, pp 1050-4729, 2014.
6. Liu, Guangyu, and Sing Kiong Nguang. "Best pole vaulting using non-causal dynamics." IEEE International Conference on Control and Automation, pp. 2323-2328 , 2009.
7. Morlier, Julien, and Mariano Cid. "Three-dimensional analysis of the angular momentum of a pole-vaulter." Journal of Biomechanics, vol 29, issue. 8, 2006
8. Nishikawa, Satoshi, Toshihiko Fukushima, and Yasuo Kuniyoshi. "Effective timing of swing-up motion by a pole-vaulting robot." International Conference on Advanced Robotics (ICAR), vol 3, pp. 1-6 , 2013.
9. [9] Pablo E., Muniz "Detection of Launch Frame in Long Jump Videos Using Computer Vision and Discreet Computation", Massachusetts Institute of Technology, 2019
10. Schade, Falk, and Adamantios Arampatzis. "Influence of pole plant time on the performance of a special jump and plant exercise in the pole vault." Journal of biomechanics vol 45, issue. 9, 2012
11. Banoth Thulasya Naika" A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends, and Research Directions", Computer Vision-Based Intelligent Systems: Challenges and Approaches, vol 12, issue 9, 2022
12. Simona, Fagaras Pia, and Graur Cristian. "Study Regarding the Speed Analysis on Approach at Junior III, Pole Vault Salt." Procedia-Social and Behavioral Sciences vol 197, 426-429, 2015.
13. J.K. Aggarwal" Human Motion: Modeling and Recognition of Actions and Interactions" International Symposium on 3D Data Processing, Visualization, and Transmission, 2017
14. Thomas B. Moeslund and Erik Granum" A Survey of Computer Vision-Based Human Motion Capture" Computer Vision and Image Understandig, vol 81, issue 3, 2001.
15. Yu-Gang Jiang1, Qi Dai1, Xiangyang Xue1, Wei Liu2, and Chong-Wah Ngo3 "Trajectory-Based Modeling of Human Actions with Motion Reference Points", European Conference on Computer Vision, pp 425–438, 2012
16. Reza Fuad Rachmadi†∗, Keiichi Uchimura†, and Gou Koutaki† "Video classification using a compacted dataset based on selected keyframe", IEEE Region 10 Conference (TENCON) vol 8, pp 873-878, 2016