HUMAN ACTIVITY DETECTIONUSING MACHINE LEARNING

[1]Ch.Sindhu Priyanka, B.Tech, M.Tech, Assistant Professor, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[2] Kavyanjali Gottumukkala, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[3] Ramya Chokka, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[4] Harichandra Prasad Baigani, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[5] Rajesh Pavan Bhaskarla, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

Abstract: Human activity recognition , or HAR for short ,is a broad field of study concerned with identifying the specific movement or action of a person based on sensor data. The sensor data may be remotely recorded ,such as video, radar, or other wireless methods. It contains data generated from accelerometer , gyroscope and other sensors of Smart phone to train supervised predictive models using machine learning techniques like SVM , Random forest and decision tree to generate a model. Which can be used to predict the kind of movement being carried out by the person which is divided into six categories walking, walking upstairs, walking down-stairs, sitting, standing and laying.

1. INTRODUCTION

Physical activity is well-known by the general public to be crucial for leading a healthy life. Thus, researchers are seeking a better understanding of the relationship between physical activity and health. Precise recording of the conducted activities is an essential requirement of their research. (Bauman et al., 2006) This data can be used to design and construct activity recognition systems. These systems allow physicians to check the recovery development of their patients automatically and constantly (da Costa Cachucho et al., 2011). Another rising concern is the sedentary life many people live, due to the shift in lifestyle occurring in the modern world, where work and leisure tend to be less physically demanding (Gyllensten, 2010). Several reports have already found links between common diseases and physical inactivity (Preece et al., 2009). Thus, activity recognition can be used by recommender systems to help the users track their daily physical activity and promote them to increase their activity level. With the recent progress in wearable technology, unobtrusive and mobile activity recognition has become reasonable. With this technology, devices like smartphones and smartwatches are widely available, hosting a wide range of built-in sensors, at the same time, providing a large amount of computation power. Overall, the technological tools exist to develop a mobile, unobtrusive and accurate physical activity recognition system. Therefore, the realization of recognizing the individuals' physical activities while performing their daily routine has become feasible. So far, no-one has investigated the usage of light-weight devices for recognizing human activities. An activity recognition system poses several main requirements. First, it should recognize activities in real-time. This demands that the features used for classification are computable in real-time. Moreover, short window durations must be employed to avoid delayed response. Finally, the classification schemes should be simple, light-weight and computationally inexpensive to be able to run on hand-held devices.

2. LITERATURE SURVEY

2.1 H.Wang,A .Klaser,C.Schmid and C-L.Liu, "Action recognition by dense trajectories ," in Proc. IEEE Conf. Comput. Vis.Pattern Recog.,Jun. 2011,pp 3169-3176.

Feature trajectories have shown to be efficient for representing videos. Typically, they are extracted using the KLT tracker or matching SIFT descriptors between frames. However, the quality as well as quantity of these trajectories is often not sufficient. Inspired by the recent success of dense sampling in image classification, we propose an approach to describe videos by dense trajectories. We sample dense points from each frame and track them based on displacement information from a dense optical flow field. Given a state-of-the-art optical flow algorithm, our trajectories are robust to fast irregular motions as well as shot boundaries. Additionally, dense trajectories cover the motion information in videos well. We, also, investigate how to design descriptors to encode the trajectory information. We introduce a novel descriptor based on motion boundary histograms, which is robust to camera motion. This descriptor consistently outperforms other state-of-the-art descriptors, in particular in uncontrolled realistic videos. We evaluate our video description in the context of action classification with a bag-of-features approach. Experimental results show a significant improvement over the state of the art on four datasets of varying difficulty, i.e. KTH, YouTube, Hollywood2 and UCF sports.

2.2 H.Wang and C Schmiid , "Action recognition with improved trajectories," in Proc.IEEE Int. Conf.Comput. Vis., Dec 2013,,pp 3551- 3558.

Recently dense trajectories were shown to be an efficient video representation for action recognition and achieved state-of-the-art results on a variety of datasets. This paper improves their performance by taking into account camera motion to correct them. To estimate camera motion, we match feature points between frames using SURF descriptors and dense optical flow, which are shown to be complementary. These matches are, then, used to robustly estimate a homography with RANSAC. Human motion is in general different from camera motion and generates inconsistent matches. To improve the estimation, a human detector is employed to remove these matches. Given the estimated camera motion, we remove trajectories consistent with it. We also use this estimation to cancel out camera motion from the optical flow. This significantly improves motion-based descriptors, such as HOF and MBH. Experimental results on four challenging action datasets (i.e., Hollywood2, HMDB51, Olympic Sports and UCF50) significantly outperform the current state of the art.

2.3 Y-G Jiang,Q.Dai,X.Xue,W.Liu and C-W Ngo. "Trajectory-based modeling of human actions with motion reference points," inProc. Eur.Conf .Comput.Vis.,Oct 2012,Vol 7576,pp.425-438.

Human action recognition in videos is a challenging problem with wide applications. State of-the-art approaches often adopt the popular bag-of-features representation based on isolated local patches or temporal patch trajectories, where motion patterns like object relationships are mostly discarded. This paper proposes a simple representation specifically aimed at the modeling of such motion relationships. We adopt global and local reference points to characterize motion information, so that the final representation can be robust to camera movement. Our approach operates on top of visual codewords derived from local patch trajectories, and therefore does not require accurate foreground-background separation, which is typically a necessary step to model object relationships. Through an extensive experimental evaluation, we show that the proposed representation offers very competitive performance on challenging benchmark datasets, and combining it with the bag of-features representation leads to substantial improvement. On Hollywood2, Olympic Sports, and HMDB51 datasets, we obtain 59.5%, 80.6% and 40.7% respectively, which are the best reported results to date.

2.4 M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A spatiotemporal maximum average correlation height filter for action recognition,"in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008,pp. 1–8.

In this paper we introduce a template-based method for recognizing human actions called action MACH. Our approach is based on a maximum average correlation height (MACH) filter. A common limitation of template-based methods is their inability to generate a single template using a collection of examples. MACH is capable of capturing intra-class variability by synthesizing a single Action MACH filter for a given action class. We generalize the traditional MACH filter to video (3D spatiotemporal volume), and vector valued data. By analyzing the response of the filter in the frequency domain, we avoid the high computational cost commonly incurred in template-based approaches. Vector valued data is analyzed using the Clifford Fourier transform, a generalization of the Fourier transform intended for both scalar and vector-valued data. Finally, we perform an extensive set of experiments and compare our method with some of the most recent approaches in the field by using publicly available datasets, and two new annotated human action datasets which include actions performed in classic feature films and sports broadcast television.

2.5 [5] Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, "Action detectionin complex scenes with spatial and temporal ambiguities," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Sep.–Oct. 2009, pp.128–135. Approach to detecting human actions in videos with complex backgrounds. They tackle the challenge of spatial and temporal ambiguities in action detection by introducing a multiple instance learning framework called SMILE-SVM. This method combines motion and appearance features to improve detection accuracy, even in cluttered scenes with occlusions and varying human appearances. Their results show promising performance in real-world scenarios, such as monitoring customer behavior in shopping malls, demonstrating the effectiveness of their approach in practical applications.

2.6 [6] T. Lan, Y. Wang, and G. Mori, “Discriminative figure-centric models for joint action localization and recognition,” in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2003 210.

They developed an algorithm for action recognition and localization in videos. The algorithm uses a figure-centric visual word representation. Different from previous approaches it does not require reliable human detection and tracking as input. Instead, the person location is treated as a latent variable that is inferred simultaneously with action recognition. A spatial model for an action is learned in a discriminative fashion under a figure-centric representation. Temporal smoothness over video sequences is also enforced. We present results on the UCF-Sports dataset, verifying the effectiveness of our model in situations where detection and tracking of individuals is challenging.

3. EXISTING SYSTEM

Several investigations have considered the use of widely available mobile devices. Ravi et. al. collected data from only two users wearing a single accelerometer-based device and then transmitted this data to the phone carried by the user (Ravi et al.,2005). Lester et. al. used accelerometer data from a small set of users along with audio and barometric sensor data to recognize eight daily activities (Lesteret al., 2006). However, the data was generated using distinct accelerometer-based devices worn by the user and then sent to the phone for storage. Some studies took advantage of the sensors incorporated into the phones themselves. Yang developed an activity recognition system using a smart-phone to distinguish between various activities (Yang, 2009). However, stair climbing was not considered and their system was trained and tested using data from only four users. Brezmes et. al. developed a real-time system for recognizing six user activities (Brezmeset al., 2009). In their system, an activity recognition model is trained for each user, i.e., there is no universal model that can be applied to new users for whom no training data exists. Bayat et al. gathered acceleration data from only four participants, performing six activities. (Bayat et al., 2014) Shoaib et al. evaluated different classifiers by collecting data of smart-phone accelerometer, gyroscope, and magnetometer for four subjects, performing six activities. (Shoaib et al., 2013).

DISADVANTAGES:

1.Require the optical sensors to be attached on the human and also demand the need of multiple camera settings.

2.Wearable devices cost are high.

Algorithm: Marker based motion Capture (MoCap) Framework.

4. PROPOSED SYSTEM

The purpose of being able to classify what activity a person is undergoing at a given time is to allow computers to provide assistance and guidance to a person prior to or while undertaking a task. The difficulty lies in how diverse our movements are as we perform our day-to-day tasks. There have been many attempts to use the various machine learning algorithms to accurately classify a person’s activity, so much so that Google have created an Activity Recognition API for developers to embed into their creation of mobile applications.

ADVANTAGES:

1.We use CNN to recognize human activities and action recognition from kinetics dataset.

2.We use transfer learning to get deep image features and trained machine learning classifiers.

3.Does not require the user to carry any devices or to attach any sensors on the human

Algorithm: Convolutional Neural Networks(CNN), VGG1

SYSTEM ARCHITECTURE

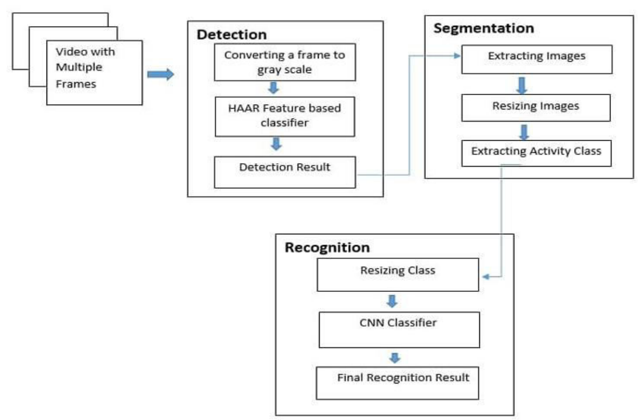

Fig 1: System Architecture

5. UML DIAGRAMS

1. CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. It is also known as a structural diagram. Class diagram contains • Classes • Interfaces • Dependency, generalization and association.

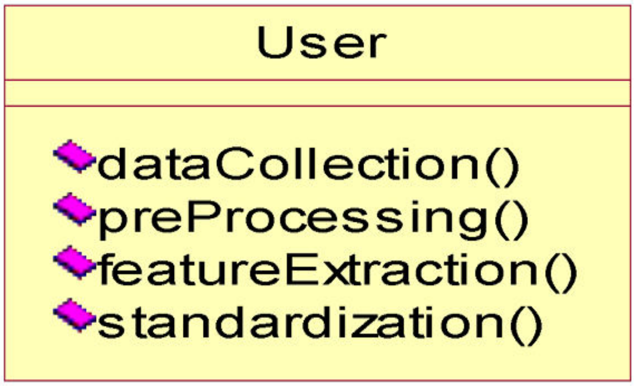

Fig 5.1 shows the class diagram of the project

2. USECASE DIAGRAM:

Use Case Diagrams are used to depict the functionality of a system or a part of a system. They are widely used to illustrate the functional requirements of the system and its interaction with external agents (actors). In brief, the purposes of use case diagrams can be said to be as follows

• Used to gather the requirements of a system.

• Used to get an outside view of a system.

• Identify the external and internal factors influencing the system.

Use case diagrams commonly contains

• Use cases

• Actors

• Dependency, generalization and association relationships.

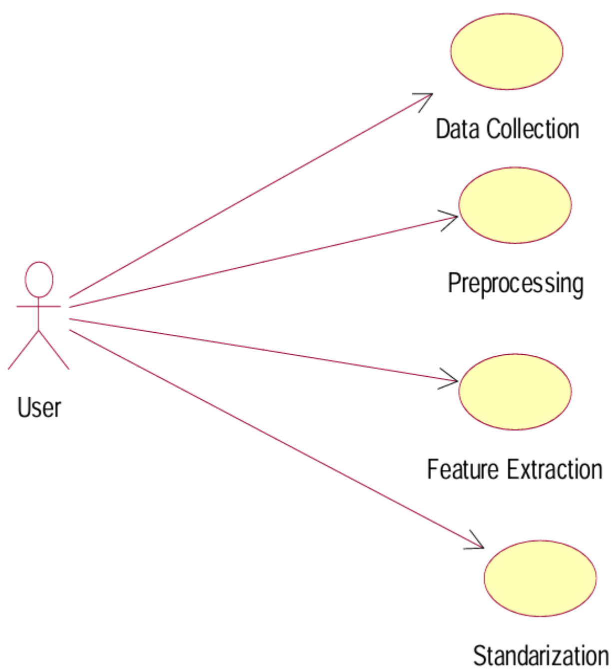

Fig 5.2 Shows the Use case Diagram

3. SEQUENCE DIAGRAM:

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. Sequence diagrams are used to formalize the behaviour of the system and to visualize the communication among objects. These are useful for identifying additional objects that participate in the use cases. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.

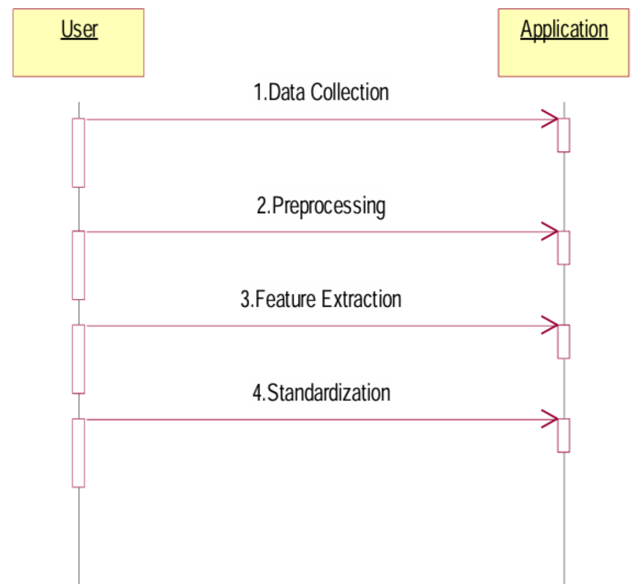

Fig 5.3 Shows the Sequence Diagram

6. RESULTS

6.1 Output Screens

In below screen click on ‘Upload Tweets Dataset’ button and upload dataset

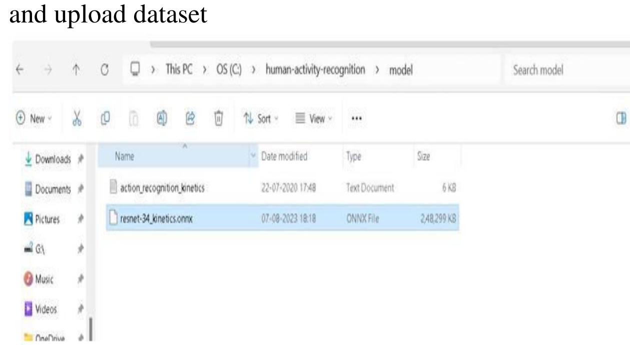

Fig 6.1 Right click on resnet file and copy its path

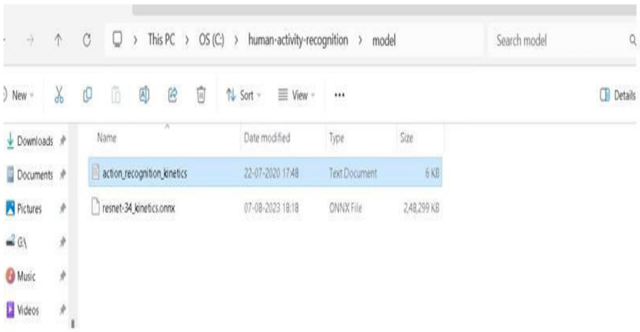

Fig 6.2 Right click on action_recognition_kinetics to copy its path

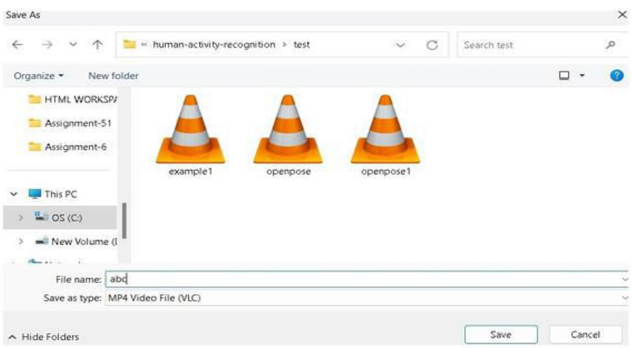

Fig 6.3 Save the video extracted from camera into system and give its path in program

Fig 6.4 The above screen shows the human action

Fig 6.5 The above Screen shows the human action and displays it in text foramt

Fig 6.6 The above Screen shows the human action and displays it in text foramt

7. CONCLUSION

In this paper, a platform to combine sensors of smartphones and smartwatches to classify various human activities was proposed. It recognizes activities in real-time Moreover, this approach is light weight, computationally inexpensive, and able to run on handheld devices. The results showed that there is no clear winner, but naive Bayes performs best in our experiment in both the classification accuracy and efficiency. The overall accuracy lies between 84.6% and 89.4%, at which the differences are negligible. Thus, this platform is able to recognize various human activities. However, all of the tested classifiers confused walking and using the stairs activities. The second conclusion is that adding the smartwatch's sensor data to the recognition system improves it's accuracy with at least six percentage point. Finally, it is computations that the best sampling frequency is in the field of 10 Hz. Some questions still require to be answered. Most important is the conducting of larger experiments with more people in order to perform more robust evaluation to clearify if indeed one method is better than the other, or whether, any off-the-shelf method can do well in this classification task. This work could be furhter extended by incorporating more sensors (e.g. heart rate sensor), recognizing high-level activities (e.g. shopping or eating dinner) or extrapolating these trained classifiers to other people.

FUTURE SCOPE

Future research in Human Activity Recognition (HAR) using Convolutional Neural Networks (CNN) can focus on several promising areas to enhance performance and applicability: 1. Data Efficiency: Developing methods to reduce the amount of labeled data required for training CNNs, such as semi-supervised learning or transfer learning techniques. 2. Sensor Fusion: Integrating data from multiple types of sensors to improve the robustness and accuracy of activity detection. 3. Temporal Dynamics: Improving the understanding of temporal dynamics in human activities by using recurrent layers or attention mechanisms alongside CNNs. 4.Energy Efficiency: Creating more energy-efficient CNN architectures that can run on wearable devices for longer periods without compromising performance. 5. Privacy Preservation: Addressing privacy concerns by developing on-device processing techniques that do not require data to be sent to the cloud. 6. Real-time Processing: Enhancing real-time processing capabilities to provide immediate feedback or intervention when necessary. 7. Complex Activity Recognition: Expanding the scope of HAR to recognize more complex activities involving multiple people or interactions with objects.

8. REFERENCES

[1] H.Wang,A .Klaser,C.Schmid and C-L.Liu, “Action recognition by dense trajectories ,” in Proc. IEEE Conf. Comput. Vis.Pattern Recog.,Jun. 2011,pp 3169-3176.

[2] H.Wang and C Schmiid , “Action recognition with improved trajectories,” in Proc.IEEE Int. Conf.Comput. Vis., Dec 2013,,pp 3551- 3558. [3] Y-G Jiang,Q.Dai,X.Xue,W.Liu and C-W Ngo. “Trajectory-based modeling of human actions with motion reference points,” inProc. Eur.Conf .Comput.Vis.,Oct 2012,Vol 7576,pp.425-438.

[4] M. D. Rodriguez, J. Ahmed, and M. Shah, “Action MACH: A spatiotemporal maximum average correlation height filter for action recognition,”in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008,pp. 1–8.

[5] Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, “Action detectionin complex scenes with spatial and temporal ambiguities,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Sep. Oct. 2009, pp.128–135.

[6] T. Lan, Y. Wang, and G. Mori, “Discriminative figure-centric models for joint action localization and recognition,” in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2003–210.

[7] M.Raptis, I.Kokkinos and S.Soatto,” Discovering discriminative action parts from mid-level video representations” ,in Proc ,IEEE Conf.Comput.Vis,.Pattern Recog.,Jun 2012,pp.1242-1249.

[8] W. Brendel and S. Todorovic, “Learning spatiotemporal graphs of human activities,” in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011 54

[9] M.Jain,J.van Genmert,H.Jegou ,P.Bouthemy and C.Snoek ,”Action localization with tubelets from motion” ,in Proc IEEE Conf, Comput.Vis.Pattern Recog. Jun 2014 pp 740-747.

[10] Sukor A.S.A., Zakaria A., Rahim N.A. Activity recognition using accelerometer sensor and machine learning classifiers; Proceedings of the 2018; Batu Feringghi, Malaysia. 9–10 March 2018; pp. 233–238