Video based human activity recognition for surveillance application

## **Introduction**

Human activity recognition attempt automatically analysis and recognize human- activities of subject with the environmental conditions using acquired information from the various types of sensors.

Human activity recognition has a hot scientific topic in computer vision community.it is involved in the development of many applications such as human computer interaction (HCI), virtual realities (VR), video surveillance for outdoor and indoor activities, home monitoring, video game, medical environment, patient monitoring, abnormal/normal activity identification, health care, and elderly people/children monitoring in specific area. HAR becomes popular and important domain in artificial intelligence and computer vision field.

Human activity has an inherent hierarchical structure that indicates the differentlevels of it. Gestures, actions, interaction, group-activities, which can be considered as a four levels categorization. First: for the bottom level, there is an “gesture” that are simple motion of a part of body. for instance, “raising an arm and moving a leg”. After action primitive level, the action/activity comes as the second level, for example: strolling, waving, walking. after that, “interaction” are human activities that involves at least two persons or objects, as an example, somebody utilize laptop, two persons checking hands.it is human to objects and human to human interaction. finally, group activities are that activities played by group composed of individuals or objects. Such as group of people playing basketball, fighting by two groups people.

In the resent years, recognizing human activities from a series of video frame or still image is a challenging task due to problems, such as human behavior, Intra-class variability, inter-class similarity, illumination changes, shadow effect, occlusion, camera jitters, noise image or frames in video, background clutter (moving background object or human) and view point etc. currently, there are limited intelligent human activity recognition system which can robustly and efficiently recognition each class of human activity. Based on the challenging and necessary of computer vision domain, HAR should be made more attention in survey and development.

Over the years, with the development of depth sensors, new opportunities arise to improve and advance this field. Recent advancement in the depth sensor (i.e., Microsoft Kinect) has made it feasible to capture the depth images/videos in real-time as well as colour images. Moreover, depth images have appropriate resolution (i.e., 640×480) and accuracy with respect to specific subject distance. As compare to the RGB images, depth images have several advantages for the problem of activity recognition. For example, depth images show appropriate geometry and shape information, which can be more discriminative than the RGB images in many problems such as segmentation, object detection and activity recognition. Depth images also provide additional body shape and structure information, which have been successfully applied to recover skeleton joints from a single depth image. Furthermore, depth images are insensitive to illumination changes. In this context, it seems very natural and important to use depth images/videos for computer vision problems. In the recent researches, it is shown that the approaches based on traditional RGB sequences could not perform well. On the other hand, depth and colour images have much different properties. The feature extraction method based on brightness, gradient and optical flow works very well for the RGB images but may not perform well for the depth images. It is therefore important to design such feature extraction method which is based on specific characteristics of depth sequences.

Briefly, the recognition of human activity involves input information from sensors, preprocessing, detection and segmentation of objects, feature extraction and representation, and classification

Related work (review of literature)

Most of the proposed techniques of video based human activity recognition was divided four categories: information resource, pro-processing, feature extraction and representation, and classification.

(L. Minh Dang a, 2020) A comprehensive survey of state-of-the-art methods, along with their pros and cons for vision-based HAR and sensor-based HAR has been provided. introduction a classification of HAR methodologies and show advantages and weaknesses for methods in each category. (Michalis Vrigkas1, 2015) in the paper, provide a detailed review of recent and state-of-the-art research advances in the field of human activity classification. And discuss their advantages and limitations of human activity methodologies. We provide a comprehensive analysis of the existing, publicly available human activity classification datasets and examine the requirements for an ideal human activity recognition dataset. (Shugang Zhang, 2017) this review highlights the advances and limited of state-of-the-art activity recognition approaches, especially for the activity representations and classification methods. And representative and available datasets are introduced. (Allah Bux Sargano 1, 2017) this review paper presents a comprehensive survey of both hand-crafted and learning-based action representations, offering comparison, analysis, and discussions on these approaches. (Agrawal1, 2017)this paper, we present the state-of-the-art which demonstrates the overall progress of suspicious activity recognition from the surveillance videos in the last decade. We include a brief introduction of the suspicious human activity recognition with its issues and challenge. (Chhavi Dhiman a, 2019)this paper aims to summarize various existing abnormal human activity recognition (Ab-HAR) hand-crafted and deep approaches with the variation of the type of information available such as two-dimensional or three-dimensional data. (Ashwin Geet D’Sa, 2019)through this paper, approaches involved in activity recognition, advantage and challenges involved in activity recognition are discussed. (Djamila Romaissa Beddiar1, 2020) this paper attempts to review and summarize the progress of HAR systems from the computer vision perspective. Most computer vision applications, challenges are discussed (Thakkar1, 2020)in this survey, we discussed various ML and DL techniques for HAR for the years 2011-2019.the paper discusses the characteristics of public datasets used for HAR.it also presents a survey of various action recognition techniques along with the HAR applications. The advantages disadvantages, and challenges of action representation, dimensionality reduction and action analysis methods are also provided.

(Ahmad Jalal1, 2015) In this paper, a video-based novel approach for human activity recognition is presented using robust hybrid features and embedded Hidden Markov Models, In the proposed HAR framework, depth maps are analyzed by temporal motion identification method to segment human silhouettes from noisy background and compute depth silhouette area for each activity to track human movements in a scene. Several representative features, including invariant, depth sequential silhouettes and Spatial-Temporal body joints features were fused together to explore gradient orientation change, intensity differentiation, temporal variation and local motion of specific body parts. (Zayed, 2015) In this paper, a system for human activity recognition is proposed. We have considered the task of obtaining a descriptive labeling of the activities being performed through labeling human sub-activities. The activities we consider happen over a long period, and comprise several sub-activities performed in a sequence. The proposed activity descriptor makes the activity recognition problem viewed as a sequence classification problem. The proposed system employs Hidden Markov Models (HMMs) to recognize human activities. (Ahmad Jalal a, 2016) In this paper, propose novel multi-fused features for on-line human activity recognition (HAR) system that recognizes human activities from continuous sequences of depth map. The proposed on-line HAR system segments human depth silhouettes using temporal human motion information as well as it obtains human skeleton joints using Spatial-temporal human body information. Then, it extracts the Spatial-temporal multi-fused features that concatenate four skeleton joint features and one body shape feature. (Youssef Hbali1, 2017) the authors propose a new skeleton-based approach to describe the Spatial-temporal aspects of a human activity sequence, using the Minkowski and cosine distances between the 3D joints. (Boufama, 2017) This paper proposes a new method using trajectories, as mid-level features, for human activity recognition. (Ahmad Jalal, 2019) In this paper, we have made an effort to allow machines understand the behavior in outer environment by proposing a novel methodology to recognize human interactions. (MUHAMMAD HAMEED SIDDIQI 1, 2019) The proposed method is an extension of the max-relevance and min-redundancy method. The ability of this method is to combine the strengths of different extraction techniques. we exploit the curve transform for feature extraction, and linear discriminant analysis for reduction of feature space. we use the hidden Markov model for activity recognition based on our proposed method of feature selection. Finally, by using the benchmark datasets such as KTH and Weizmann datasets compare the proposed scheme with state-of-the-art. Simulation results show that the proposed scheme is not only more accurate for some datasets, but outperforms competing method by weighted average accuracy 98%.

In classification. (D.K. Vishwakarma ⇑, 2015) The aim of this paper is to present a new approach for human activity recognition in a video sequence by exploiting the key poses of the human silhouettes, and constructing a new classification model. The proposed hybrid classification model is a combination of SVM and 1-NN model and termed as ‘SVM–NN’. (Uddin, 2016) A novel approach is proposed here for depth video based human activity recognition, using joint-based Spatial-temporal features of depth body shapes and hidden Markov models. From depth video, different body parts of human activities are first segmented using a trained random forest. The activity features are then further enhanced using generalized discriminant analysis to classify them nonlinearly in order to convert them to more robust features. (S. U. Parka, 2016)In this paper, we propose a new HAR system via Recurrent Neural Network (RNN) which is one of deep learning algorithms. We utilize joint angles from multiple body joints changing in time which are represented a Spatial-temporal feature matrix (i.e., multiple body joint angles in time). With these derived features, we train and test our RNN for HAR. Our test results show that the proposed RNN-based HAR is able to recognize twelve human activities reliably and outperforms the HMM- and DBN-based HAR. We have achieved the average recognition accuracy of 99.55% for the activities. The results are 7.06% more accurate than that of the HMM-based HAR and 2.01% more accurate than that of the DBN-based HAR. (Mohanad Babiker1, 2017)A robust neural network was built based on the human activities features database, which was extracted from the frame sequences. multi-layer feed forward perception network used to classify the activities model in the dataset. The classification results show a high performance in all of the stages of training, testing and validation. (Rajat Khurana, 2018) One of the most pioneering technique for Human Activity Recognition is based upon deep learning and this paper focuses on various approaches based on that. Convolution Neural Network and Recurrent Neural Networks are mostly used in deep learning architectures. (Sumaira Ghazal1, 2019)Supervised machine learning was implemented for recognizing four activity classes including sit, stand, walk and fall. Performance of five techniques including K-nearest neighbours (KNNs), support vector machine, Naive-Bayes, linear discriminant and feed-forward back- propagation neural network was compared to find the best classifier for the proposed method. All techniques performed well with best results obtained through the KNN classifier.

Finally, (Vishwakarma, 2019) categorized these datasets into two part, the first part consists two-dimensional (2D-RGB) datasets and the second part has three-dimensional (3D-RGB) datasets. presented the state-of-the-art algorithms that give the highest accuracy on these datasets

Propose approach

the recognition of human activity (HAR) involves input information from sensors/visual, pre-processing, detection and segmentation of objects, feature extraction and representation, and classification steps.

Sensor technology has achieved exceptional developments in multiple perspectives. The fast development of smart-phone, wearable devices, and CCTV systems has motivated researchers to improve HAR systems under practical situations. Now days, the depth sensor has made it feasible to capture the depth images/videos in real-time as well as colour images. The several data source such as RGB, RGB D, and various wearable devices signal from different sensor or visual used in human activity recognition systems.

Background subtraction, feature detecting and tracking, feature extraction is an important step. Skin-colour based, shape based, pixel values based, 3D models based, motion based, anisotropic-diffusion based method that various methods is obtained for detecting. Template-based methods such as features tracking, contours based tracking, or optimal estimation, particle filters, Cam shift can be sued to track the action.in (Kimia, 2016)a commercial depth-sensor-based camera, is utilized in this work to acquire the RGB as well as the depth images of different human activities. The depth silhouette is then extracted from each depth image after a background subtraction. After obtaining a depth silhouette, a corresponding skeleton-body model that provides 15 joint positions is obtained through the OpenNI library, therefore, from each depth silhouette, 15 body joints are obtained for the calculation of the spatial-temporal features. Therefore, proper pre-processing operating that include foreground/background frame segmentation for action feature detection, feature tracking, is huge magnitude for feature extraction and representation.

Finally, the features of action/activity are performed by classification methods. (Djamila Romaissa Beddiar1, 2020) For instance, Support Vector machine (SVM), Naïve Bayesian classifier, algorithm of K-Nearest neighbour, K-mean, Mean shift clustering, Machines finite state, Hidden Markov Model, Dynamic time warping, Neural networks (CNN, DNN, RNN), and hybrid method to classify the feature of Human activity.

references

1.L. Minh Dang a , Kyungbok Min a , Hanxiang Wang a , Md. Jalil Piran a , Cheol Hee Lee b , Hyeonjoon Moon(2020) Sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Recognition 108 107561

2. Vrigkas, Michalis, Nikou, Christophoros, Kakadiaris, Ioannis A. (2015) A Review of Human Activity Recognition Methods. Frontiers in Robotics and AI 2 2296-9144

3. Shugang Zhang,1 Zhiqiang Wei,1 Jie Nie,2 Lei Huang,1 Shuang Wang,1 and Zhen Li1. (2017) A Review on Human Activity Recognition Using Vision-Based Method. Hindawi Journal of Healthcare Engineering. 2040-2309

4. Sargano, Allah Bux, Angelov, Plamen, Habib, Zulfiqar (2017) A Comprehensive Review on Handcrafted and Learning-Based Action Representation Approaches for Human Activity Recognition. Applied Sciences (Switzerland). 7 2076-3417

5. Dhiman, Chhavi Vishwakarma, Dinesh Kumar (2019) A review of state-of-the-art techniques for abnormal human activity recognition Engineering Applications of Artificial Intelligence 09521976 21-45

6. Ashwin Geet D’Sa Dr. B G Prasad (2019) A survey on vision-based activity recognition, its applications and challenges 2019 2nd International Conference on Advanced Computational and Communication Paradigms, ICACCP 2019 9781538679890

7. Djamila Romaissa Beddiar1,3 · Brahim Nini1 · Mohammad Sabokrou2 · Abdenour Hadid3(2020) Vision-based human activity recognition: a survey. Multimedia Tools and Applications 15737721 41-42

8. Preksha Pareek1 · Ankit Thakkar1 (2020) A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications Artificial Intelligence Review 15737462

9. Ahmad Jalal1, Shaharyar Kamal2 and Daijin Kim1(2015) Depth silhouettes context: A new robust feature for human tracking and activity recognition based on embedded HMMs 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI 2015 9781467379700

10. Zayed, Ahmed Taha Hala, El-Sayed M. El-Horbarty (2015) Skeleton-based Human Activity Recognition for Video Surveillance, International Journal of Scientific & Engineering Research 2229-5518 993-1004

11. Jalal, Ahmad Kim, Yeon Ho Kim, Yong Joong Kamal, Shaharyar Kim, Daijin (2016) Robust human activity recognition from depth video using spatiotemporal multi-fused features Pattern Recognition 0031-3203

12. Youssef Hbali1 , Sara Hbali1, Lahoucine Ballihi2, Mohammed Sadgal1 (2017) Skeleton-based human activity recognition for elderly monitoring systems IET Computer Vision 1751-9640 16-26

13. Jalal, Ahmad Mahmood, Maria Hasan, Abdul S. (2019) Multi-features descriptors for Human Activity Tracking and Recognition in Indoor-Outdoor Environments Proceedings of 2019 16th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2019 9781538677292 371-376

14. MUHAMMAD HAMEED SIDDIQI 1, MADALLAH ALRUWAILI 1, AND AMJAD ALI 2(2019) A Novel Feature Selection Method for Video-Based Human Activity Recognition Systems IEEE Access 21693536 119593-119602

15. D.K. Vishwakarma ⇑, Rajiv Kapoor (2015) Hybrid classifier based human activity recognition using the silhouette and cells Expert Systems with Applications 09574174 6957-6965

16. Uddin, Md Zia (2016) Human activity recognition using segmented body part and body joint features with hidden Markov models. Multimedia Tools and Applications 15737721 13585-13614

17. Mohanad Babiker1, Othman O. khalifa1, Kyaw Kyaw Htike2, Aisha Hassan1, Muhamed Zaharadeen1 (2017) 2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017 9781538639603 1-5

18. RajatKushwaha, Alok Kumar Singh(2018) Deep Learning Approaches for Human Activity Recognition in Video Surveillance - A Survey ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications 9781538663738 542-544

19. Md. Zia Uddin and Jaehyoun Kim(2016) Human activity recognition using spatiotemporal 3-D body joint features with hidden markov models KSII Transactions on Internet and Information Systems 2288-1468 2767-2780