International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228392

Survey On Feature Extraction Approach for Human Action Recognition in Still Images and Videos

Pavan M[1], Deepika D[2], Divyashree R[2], Kavana K[2], Pooja V Biligi[2]
[1]Assistant Professor, Department of ISE, JNNCE, Shivamogga, Karnataka, India
[2]Department of ISE, JNNCE, Shivamogga, Karnataka, India

Article Info
Volume 8, Issue 3
Page Number : 359-369

Publication Issue :
May-June-2022

Article History
Accepted: 05 June 2022
Published: 20 June 2022

ABSTRACT

Human Action Recognition (HAR) has been a challenging problem yet it needs to be solved. Recently the detection and recognition of human action has broad range of applications and is popularized in the field of computer vision. It mainly focuses on to understand human behaviour and name a label to each action. There are many approaches for action recognizing from both image and video based actions. Now it is time to review these existing approaches in order to help for future research. The main aim of this work is to study the various action recognition techniques in videos and images. The paper presents a brief overview of features of human actions by categorizing as still image-based and video-based. All related datasets are also introduced in this paper, which will be helpful for future research.

Keywords : Computer vision; Human Action Recognition; Still image-based; Video-based.

I. INTRODUCTION

With the rapid increase in evolution of technologies, many new applications are built to monitor and assess the action and response of human being. These are not only limited to motion of some particular muscles but also for different systematic behaviors performed. This type of action recognition will provide information about humans and it will help us to appreciate their feelings, intentions and frame of mind of them.

Human Action Recognition is presently one of the most demanding applications in worldwide. Computer vision is a sub branch of Artificial Intelligence (AI) which will help to extract significant information from images or videos and further use of this information to take actions. If Artificial Intelligence make the computer to think, computer vision will make the system to observe and understand. Computer vision specially deals with the explanation, understanding and recreation of a 3D scene from videos and images[1]. Nowadays computer vision has becoming increasingly popular and it is used to solve many problems. One of such problem in computer vision is Human Action Recognition (HAR). HAR is the ability of computer to identify, analyze and understand human behavior and actions to predict the output similar to the way of the humans

do. Here actions refer to the movement of body parts. These actions can be either atomic or primitive actions. Atomic refers to the actions performed by specific body parts like hand, leg, upper body part, etc. Primitive actions are executed in successive order and these are set of atomic actions such as "raising right hand" and "stretching right arm". Recently Human Action Recognition becomes popular in the field of computer vision, image processing and machine learning.

Nowadays, there is increase in demand for human action recognition due to many applications like human-computer interactions, health care, video surveillance, frame tagging, etc. HAR plays a very important role in interpersonal relations because it provide information about person identity, their psychological state and their personality which is difficult to derive[2]. The increase demand of human action recognition has drawn a lot of attention in the field of computer science due to many domain applications, for instance in security surveillance systems, healthcare systems, human-computer interaction application, searching an image using verbs, frame tagging, robotics applications and in other physical and metrological applications.

Detecting the actions of humans from video database is a challenging task. The accuracy and performance of action recognition task is based on image segmentation efficiency, finding Region of Interest (ROI), extracting image features and classification of images by labeling them. Action recognition is difficult in still images is compared to video because there will be no spatio-temporal features, high intra class variance, low inter class variance in some actions, background clutter, change in background lightening. Removing background from an image itself is more challenging because there will be no frames to take difference. In videos spatio-temporal features are important to classify the actions, but in case of images, there will be no temporal features which make it difficult to recognize the actions.

The paper is organized as follows. We present different features for action recognition in images in section 2. Various features for action recognition in videos is presented in section 3. The database that has been used for action recognition is briefed in section 4. Finally, in section 5 it includes the conclusion of the paper.

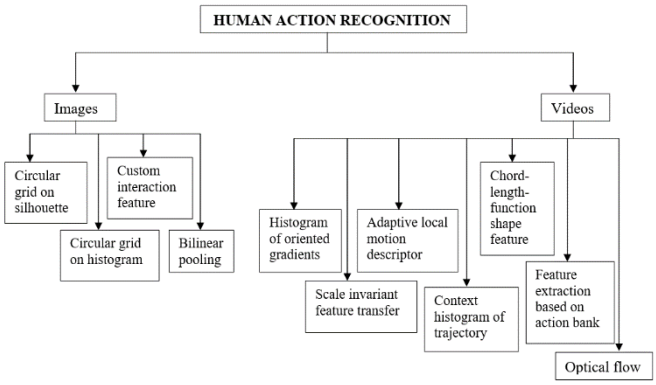

Figure 1: General overview of Human action recognition

Here is an overview of Human action recognition which consists of images and videos. Further images is described in four methods such as circular grid on silhouette, circular grid on histogram, custom interaction feature, bilinear pooling, where as videos is described into seven methods they are histogram of oriented gradients, scale invariant feature transfer, adaptive local motion descriptor, context histogram of trajectory, chord-length-function shape feature, feature extraction based on action bank, optical flow.

II. FEATURES FOR STILL IMAGES BASED ACTION RECOGNITION

2.1. Circular grid of histograms

The Conditional Random fields[3] is used to build a deformable images based on edge and region features. Position of body parts is obtained from edge-based deformable model and each body part image is represented by regional model. Circular grid is placed over human silhouette, which will help to obtain positions of all the parts. This grid consists of 12 bins which are 30 degrees apart. Using convolution of a

rectangular filter, rectangular patches are searched over human silhouette. Then, pose is represented by Circular Histogram of rectangles (CHORs). The rectangular patches are searched from the human silhouette. After finding the rectangles, CHOR is used to represent the pose. The bins of this circular histogram are 12 in number, where each bin are 30 degree apart. The entire process is depicted in figure 2. Using a rectangle histogram approach, we can discriminate actions to a great extent. This approach is also used in an unsupervised setting. The higher accuracy rate can be achieved by supervised classification on still actions dataset. However, the actions that are indistinguishable to naked eye makes actions with similar kind of poses make the pose extraction slightly difficult. Due to the absence of region model in still images, segregating foreground and background objects is hard.

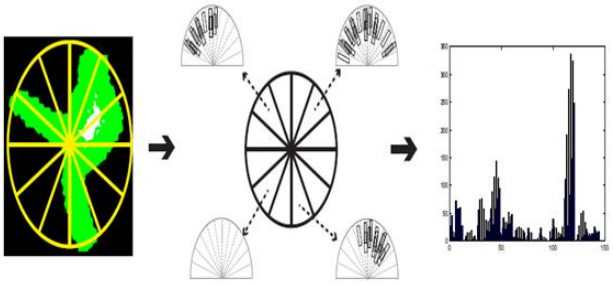

Figure 2: Representation of pose using CHORs.

2.2. Custom interaction feature

Custom interaction feature[4] is used to know the spatial relationship between 2 action components. Here, union and intersection region is computed using coordinates of person and object bounding box. Then, Euclidean distance is computed between centers of components and angle feature is eight dimensional sparse vectors, where 360 degrees is divides into eight bins. The angle present in bin is denoted by 1 and rest are represented by 0. Then the diagonal of union region is used to normalized the distance between the centers of two regions is divides into eight bins. Figure 3 represents the action recognition by concatenating person, interaction, object and union features algorithm flowchart.

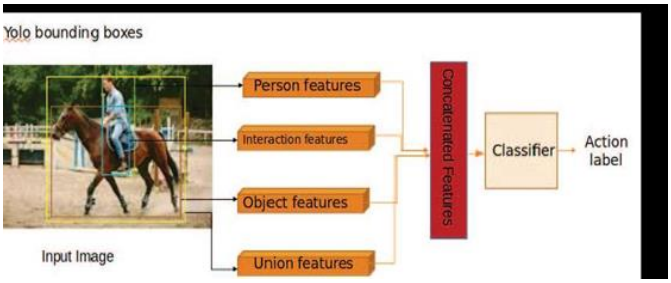

Figure 3: Action recognition algorithm flowchart

The approach of using all the components of image along with custom interaction feature will increase the accuracy of action recognition. First separate all the components from images, extract the features from each components and then process them together will make this model robust for any variations in the dataset. In some cases image with larger dimensions is less efficient as euclidean distance in larger dimensions are not much efficient.

2.3. Attention bilinear pooling with mask aggregation method

This model is divided into two important components that is channel and spatial wise attention mechanism[5]. Important channel is extracted from channel wise attention add weights are added to increase the channel information and unimportant information is suppressed. Different weights are allocated to different positions using spatial-wise attention, therefore useful areas are strengthen and weak the useless areas. Then, all the attention feature maps are combined to generate an aggregated mask to extract ROIs which is used to reduce background noise. For adjustment of different mask pooling the operation used is equal scale which are generated by attention feature because feature extracted from different layer have different dimensions.

The channel-spatial model is effective to increase accuracy of human action recognition as it can detect more discriminative regions and this method gives outstanding performance for interaction with larger

objects as it can be easily identified, whereas its accuracy decreases for smaller objects.

2.4. LLC features and GIST features

This method combines dense Scale -Invariant Feature Transform (SIFT) features encoded and pooled by Locality-constrained Linear Coding (LLC)[6] and GIST features together to describe the still image. SIFT features are extracted by dividing the image into blocks with each region of the part with size 16*16 pixel. Distance between any 2 parts is 8 pixels, then histogram of gradient direction is computed for all those parts in 8 directions, leading to a descriptor of 128 dimension for each part. GIST features are obtained by dividing the image of interest into sub blocks of image of size 4*4 (16). Gabor filter group considering 32 scales and orientation is built to bring about convolution operation on each image block by using each Gabor filter in the group. For the entire image a descriptor of 512 dimension is obtained.

Images can be described more accurately by using descriptors in combination rather than a single descriptor as shown in figure 4. However, when the image has a complex background or some parts of the body are obscured, it is not easy to accurately estimate the posture of the human body. Sometimes it can recognize the wrong posture, which will affect the final recognition effect.

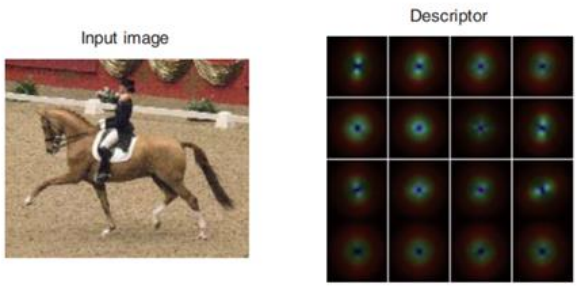

Figure 4: GIST descriptor

2.5. Human-object relation network

Human-object relation module is used to recognize actions in still images. Initially, using convolution blocks image-level feature maps are extracted, with ROI pooling[7] instance-level features are attained. Human feature fh and object feature fo obtained by instance-level features with these two as input, relation weight who is computed using scaled dot formula. Since spatial location is also an important factor we calculate geometry weight wg using human and object bounding boxes in correspondence with position embedding. All these are used to calculate fho and to calculate foh we consider transposed who. Multiple human-object relation modules are made use to compute relation-enhanced human-object and object-human features.

Especially for some kinds of actions object features are also required in addition to human features for proper recognition. It reduces the noise of unrelated information and enhances object features. It can be trained with present action recognition datasets without much hassle.

2.6. Knowledge Distillation Framework for Action Recognition in Still Images

This method considers two networks namely the teacher which is pre-trained and complex and the student network[8]. The teacher network is initially trained using Stanford 40 dataset. It uses soft max as an activation function to generate soft labels which are class probabilities with temperature(T). The student network is trained with the goal of attaining the results which are similar to that of a complex network. Loss is calculated for both the models using probabilities obtained by both the models, Kullback-Leibler divergence in between 2 probability distributions and cross-entropy loss. Back propagation is carried out only the student network. All the steps for 4 pairs of student and teacher networks are repeated. To improve the overall network performance squeeze and excitation bocks are included in few cases.

Auxiliary data at the time of training and testing is not essential in this approach. The computational costs, number of parameters to be considered are lesser as compared to deep CNN networks.

Improvements to the approach has to done by applying deeper teacher networks to attain improved performance of student network.

III. FEATURES FOR VIDEO BASED ACTION RECOGNITION

3.1. Circular grid on silhouette

In this method features are extracted based on silhouette[9]. First, N points is calculated from the silhouette and centroid is calculated using these points. Next, smallest circumference which is divided into n equal segments is placed on silhouette as shown in below figure 5. Then, arc is computed in radians in right to left direction of corresponding circumference. The Distance is calculated from centroid to closest and furthest points in each circle segment. Then at last feature vector V is obtained from all segments.

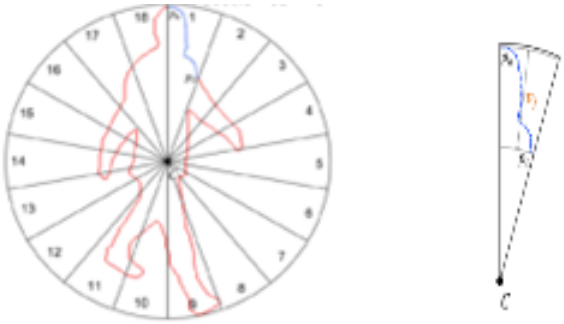

Figure 5: Silhouette inscribed by smallest circumference and each point in the silhouette has the distance with respect to centroid C is computed.

In this method feature extraction is relatively faster and performance is more stable, but it cannot be used for non static cameras.

3.2. Histogram of oriented gradients (HOG)

HOG is a method to calculate the gradient for the localized parts of an image[10]. The complete picture is separated into small parts known as cells. For each pixel in connected path, the directions of HOG are founded. All the cells which are[00, 1800] periods based on the gradient orientation are separated. Each pixel gives a weighted vote to its angular bin. The group of cells obtained gets convert into large blocks which are interlinked and the gradient strengths which is obtained and which represents the histogram for all block is normalized. Finally the descriptor is represented from the obtained histogram of all blocks.

HOG method leads to more accuracy for crowd analysis and faster processing. Minimum number of false action recognition is done.

3.3. 2D Scale invariant feature transform (SIFT)

All the image scale will be considered as scale invariant by the SIFT features[11]. In order the image produce the scale space, Gaussian function uses low as the scale-space kernel. The interested points are detected by DoG image, which have some difference of smoothed images. From the DoG images, local extremes are detected. The interest points from all the pixels of the smoothed image are calculated. The gradient orientation and magnitude in the region construct the 36 directions of weighted histograms. The interested point and the peak that is more than 80%of the histogram's maximum value is assumed as the orientation of interest point. Around the interest points region rotates the orientation, then descriptor is created. Then, the region gets divide into 4 x 4 blocks. At each block, the histogram of eight directions is constructed. Thus, the calculated feature vector for each point is 4 x 4 x 8=128.

The results are more easier to compare and simple to understand. The approach for human action recognition in this method gives more comprehensive to extract the features.

3.4. Adaptive local motion descriptor

In this method the proposed descriptor collects dynamic texture[12] traits from frames. By considering 2 successive frames and taking the average of the adjacent pixel as the threshold value, it creates a dynamic movement statistics and also a consistent pattern in stable and non-stable regions. In

order to manage intensity variations additional bits are considered. The movement feature is to begin with calculated the usage of consecutive frames (preceding body and subsequent body). The frames are then divided into small cells, and the adjoining pixels are as compared to the following body centre pixel value and threshold value, that's decided via way of means of averaging the median values, which might be calculated the usage of absolutely the distinction between the centre pixel and its close by pixels with inside the body cells. Following that, the ensuing code is break up into higher and lower patterns for each frame. The XOR (unique OR) is used to merge higher and lower patterns from the preceding and subsequent frames.

The figure 6 is an example of ALM Descriptor where the previous and next frames are considered to produce upper and lower patterns respectively and changes in the pixels between the frames is found out, later histogram features are obtained. ALMD it describes texture feature as well as the motion. It has more accuracy than the state of the art works. Drawbacks are that the actions with similar kind of poses and actions which are indistinguishable to naked eye makes pose extraction slightly difficult and due to the absence of region model in still images, segregating foreground and background objects is hard.

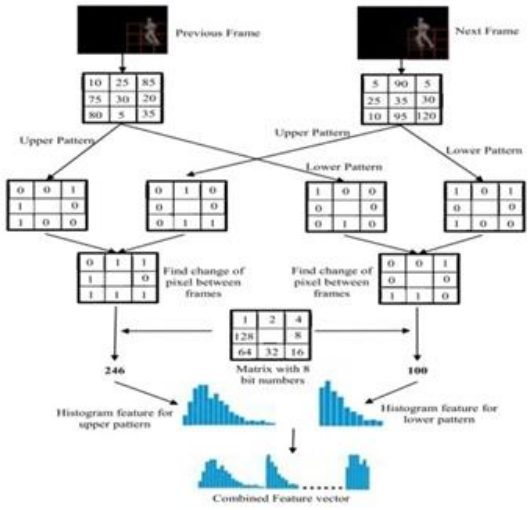

Figure 6: An example of Adaptive Local Motion Descriptor

3.5. Context Histogram of Trajectory Feature Descriptor (CHOT)

The described descriptor[13] is fabricated to transmute the line into a trajectory feature that comprises movement trajectory space information. The algorithm divides human space into various sections grounded on the space separation of shape environment as 2 * 8 subdivisions, with each subdivision further split into 2 * 8 regions. Also, to effectively distinguish the movements of distinct sections of the human, direction histograms for distinguishable regions in the human space are extracted, and a full histogram is constructed. In the following frames, It implies that the trajectory feature can be attained between every two successive frames. Finally, a frame action is used to create CHOTs with 256 dimensions. The figure 7 describes a CHOT Histogram which comprise of 16 sub histograms composed of 16 sub bins, making it total of 256 bins.

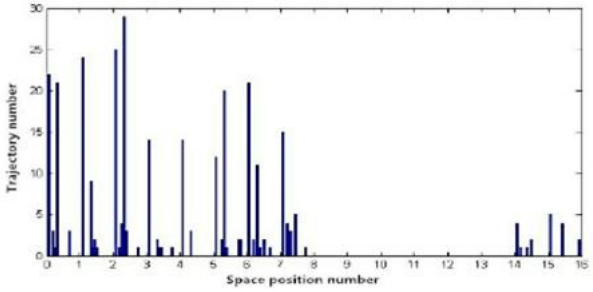

Figure 7: CHOT Histogram

CHOTs helps in reaching higher accuracy rates even when there is inconsistency and the foreground is not that absolute and it extracts the local features from images helps in increasing the overall performance. Some issues regarding the robustness and efficiency of the action recognition still need to be solved.

3.6. Feature extraction based on Action bank features

Action bank features[14] are calculated from the fixed set of videos. The predefined set of videos is called as an action bank. From templates calculation is done for similar videos. Here the action bank videos act as templates. Comparison of new video is against the

action bank video to get the action bank feature of that video. The similar information of input video with the action bank video is to be captured in order to compute73-element vector. As the result of similar information of the new video is against all the action bank videos, action bank features of the new video are generated. By using an n action bank videos, action bank feature of size n× 73 is generated. Each horizontal line in the below figure 8 corresponds to an action bank feature. Hence, videos with similar actions may contain same or close patterns in action-bank features and also contain similar local patterns, based on the extent of similarity and nature. By computing the similarity between the video and the respective action-bank video, the horizontal lines are calculated.

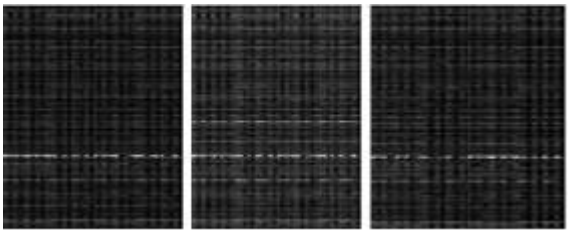

Figure 8: Features of Action bank running videos in KTH dataset

The results obtained from this approach is effective in local pattern recognition on UCF-50 dataset. This approach should be developed further in order to evaluate other features that captures the spatio-temporal distribution of information in other datasets.

3.7. Optical flow based approach

Optical flow[15] vectors builds a local descriptor along the action performer's edge. The silhouette variations velocity and time vectors along with the silhouette is captured efficiently by using optical flow based approach. The boundary is calculated by feature vector and the feature set of optical flow based includes the shape. Here, the performers are nothing but the instantaneous velocity information which is extracted along the action boundaries. Next, the centre of gravity (CG) of the foreground is calculated. Then, in order to intersect the boundary lines, from the centroid, the 5k degrees of radial lines are drawn (k is an integer) to intersect the boundary lines. These points are computed by the boundary points with radial distances which lie on these.

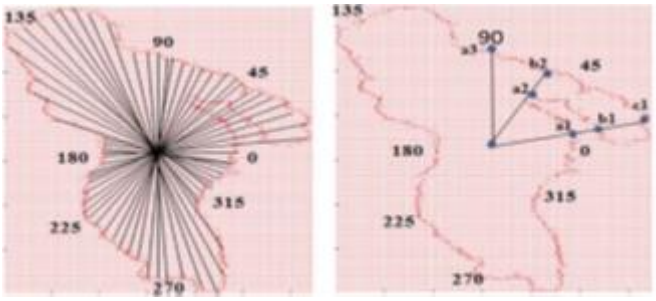

Figure 9: Optical flow based feature extraction for human actions.

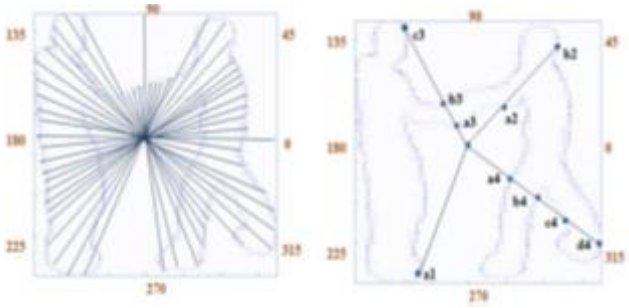

Figure 10: Optical flow based feature extraction for human interactions.

The figure 9 represents the feature extraction based on optical flow approach for human actions. Similarly figure 10 represents the optical flow-based feature extraction for human to human interactions. Based on the radial line at particular angle which intersects at particular point, the feature vector is calculated. The Optical flow based approach demonstrates that this method is simple but yet efficient. This method achieves high accuracy for Weizmann dataset and KTH dataset. Hence, this approach is considered as good option for human action as well as interaction classification for video related applications. This method is improved further to recognize the actions in other datasets

IV. DATASETS

4.1 A Multicamera Human Action Video Dataset for Action Recognition (MuHAvi)

MuHAVi-uncut[3] is a realistic dataset that includes much longer sequences with more performers and cameras. It features 17 activities performed by seven different people and captured by eight separate cameras.

4.2 KTH dataset

The KTH action dataset[3] contains various human actions such as running, walking, boxing, and waving hand. There are 25 individuals who perform these actions in different settings, such as indoors, outdoors, and with different clothes.

4.3 MS COCO dataset

The gold standard benchmark for evaluating computer vision models is the MS COCO (Microsoft Common Objects in Context) [6]. It involves the annotations like object detection, segmentation, key point detection, and captioning on a big scale. It consists of 328k images.

4.4 Stanford 40 action dataset

The Stanford 40 action dataset[5] consists of forty variety of actions with 9532 total pictures with 180-300 images per action class. Here the actions belong to various categories. The class of every action of image have giant variations in backdrop, human stance and outward form.

4.5 UT-interaction dataset

The data collected from the 20 videos[7] were captured by a static camera at 25 frames per second. The interactions in the dataset are categorized into six classes: handshakes, push, kick, punch, and point. The interaction dataset of the University of Texas at Austin consists of 10 video sequences. Each of these features an individual with over 15 different clothing. The videos were captured with a resolution of 720 x 480 and a height of around 200 pixels.

4.6 WEIZMANN Dataset

The Weizmann dataset[8] comprises of ninety videos with total variety of ten actions performed by nine individuals contains 90 videos. The actions are jumping jack, waving hand, running etc.

4.7 UCF Sports action dataset

The UCF Sports action dataset[9] embodies hundred and fifty video Sequences. It includes the set of sports actions which consists ten human actions in various sports. It includes the 150 sequence of natural pool actions with numerous applications.

4.8 UCF-50 dataset

UCF50[9] contains realistic action videos having 50 action classes. They are collected from YouTube. This dataset is quite challenging because of large variations in camera motion, viewpoint, background clutters, illumination conditions, etc. and for all 50 classes, the videos are divided into 25 groups, where each group consists of more than 4 action clips. The actions include Basketball shooting, diving, drumming, walking with a dog, pole vault, etc.

4.9 Willow action dataset

Willow action dataset[6] consists of classification of images in human actions. Large variation is detected in poses and appearances and it consists of 24 multi-view sequence with total number of 353 RGB-D frames and 7 actions.

4.10 PASCAL VOC 2012 dataset

PASCAL Visual Object Classes (VOC) 2012 dataset[7] consists of 256 classes. It includes 20 object categories like vehicles, boat, sofa, horse, train, etc. There are 3 subsets of dataset: 1464 images for training, 1449 images for validation and a private testing set.

Table 1: Dataset assembled for still images and videos for action recognition

Dataset	Type	Actions	Used in papers	Year
MuHAVi dataset	Image	17	[3], [19], [20]	2018, 2016
KTH action dataset	Image and video	6	[15], [16], [17]	2020, 2021
MS COCO dataset	Image	89	[6]	2020
Stanford 40 action dataset	Image	40	[21], [22], [23]	2020, 2021
UT-interaction dataset	Video	6	[18], [17]	2016, 2021
WEIZMANN dataset	Video	10	[29], [30]	2010, 2018
UCF Sports action dataset	Video	10	[7], [24], [26]	2017, 2018, 2020
UCF-50 dataset	Video	50	[27], [28]	2017, 2015
Willow action dataset	Image	7	[6]	2017
PASCAL VOC 2012 dataset	Image	20	[7]	2020

V. CONCLUSION

A survey of numerous approaches to video-based and image-based action recognition has been undertaken in this work. On the basis of a categorization of image-based action recognition and video-based action recognition, different feature extraction algorithms are presented on a categorization on both the methods for action recognition. Six different feature extraction approaches for image- based recognition have been explained. In circular grid of histogram, pose is extracted by considering the histogram of rectangular patches in bins of human silhouette. From the custom interaction feature method, feature is obtained by computing the union and intersection region. Important features are gained in attention bilinear pooling with mask aggregation method by considering the channel and spatial attention mechanism. By training teacher and student network in knowledge distillation method action is recognized. Under the video-based recognition category, total 7 feature methods are introduced. By considering the silhouette of the human and the circular grid, feature extraction is done by Circular grid on silhouette approach. In HOG method, histogram of blocks represents the feature descriptor. The feature extraction is done by considering all the scales of an image which is to be scale invariant in 2d SIFT. Both motion feature and the texture is extracted by ALMD. The trajectory feature description of CHOT method considers the trajectory to obtain the feature. Action bank features are extracted using action bank feature method. In this paper, all the datasets that are used by the different methods are briefly explained. Still image-based action recognition research is still in its early stages, as it is a relatively new field. We hope that this study will spark additional research efforts in the field of human action identification in still images and videos.

VI. REFERENCES

[1]. Jagadeesh B, Chandrashekar M Patil, “Video Based Action Detection and Recognition Human using Optical Flow and SVM Classifier”, In proceedings of IEEE International Conference On Recent Trends In Electronics Information Communication Technology, India, May 20- 21, 2016.

[2]. Denver Naidoo, Jules-Raymond Tapamo, Tom Walingo, “Human Action Recognition using Spatial- Temporal Analysis and Bag of Visual

Words", In proceedings of 14th International Conference on Signal- Image Technology & Internet-Based Systems (SITIS)2018.

[3]. NazliIkizler, R. GokberkCinbis, SelenPehlivan and Pinar Duygulu, "Recognizing Actions from Still Images", 2008.

[4]. Deeptha Girish, Vineeta Singh, Anca Ralescu, "Understanding action recognition in still images", In proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.

[5]. Wei Wu, Jiale Yu, "An Improved Bilinear Pooling Method for Image-Based Action Recognition", In proceedings of 25th International Conference on Pattern Recognition (ICPR), 2020.

[6]. W. Ende, H. Xukui and L. Xuepeng, "Static human behavior classification based on LLC features and GIST features," 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 2017, pp. 651-656.

[7]. W. Ma and S. Liang, "Human-Object Relation Network For Action Recognition In Still Images," 2020 IEEE International Conference on Multimedia and Expo (ICME), 2020, pp. 1-6.

[8]. M. chapariniya, S. S. Ashrafi and S. B. Shokouhi, "Knowledge Distillation Framework for Action Recognition in Still Images," 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE), 2020, pp. 274-277.

[9]. González, L., Velastin, S.A y Acuña, "Silhouette-based human action recognition with a multi-class support vector machine", In proceedings of 9th International Conference on Pattern Recognition Systems (ICPRS)2018.

[10]. Chandrashekar M Patil, Jagadeesh B, Meghana M N, "An Approach of Understanding Human Activity Recognition and Detection for Video Surveillance using HOG Descriptor and SVM Classifier", In proceedings of International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC), 2017.

[11]. Jia Liu, Jie Yang, Yi Zhang, Xiangjian He, "Action Recognition by Multiple Features and Hyper-sphere Multi- class SVM", In proceedings of International Conference on Pattern Recognition, 2010.

[12]. M. A. Uddin, J. B. Joolee, A. Alam and Y. -K. Lee, "Human Action Recognition Using Adaptive Local Motion Descriptor in Spark," in IEEE Access, vol. 5, pp. 21157- 21167, 2017.

[13]. L. Zhu, Q. Zhou and Z. Li, "A New Method of Feature Description for Human Action Recognition," In proceedings of 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2016, pp. 396-400, 2016.

[14]. SamySadek, Ayoub Al-Hamadi, Bernd Michaelis and Usama Sayed, "An SVM approach for activity recognition based on chord-length-function shape features", 2012.

[15]. Earnest Paul Ijjina, C Krishna Mohan, "Human action recognition based on recognition of linear patterns in action bank features using convolution neural networks", In proceedings of 13th International Conference on Machine Learning and Applications, 2014.

[16]. S.Santhosh Kumar and Mala John, "Human Activity Recognition using Optical Flow based Feature Set", 2016.

[17]. S. Shi and C. Jung, "Deep Metric Learning forHuman Action Recognition with Slow Fast Networks," 2021 International Conference on Visual Communications and Image Processing (VCIP), 2021, pp.1-5.

[18]. S. P. Sahoo and S. Ari, "A Three Stream Deep Network on Extracted Projected Planes for Human Action Recognition", 2020 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2020, pp. 1-5.

[19].S. S. Mohith, S. Vijay, S. V and N. Krupa, "Trajectory Based Human Action Recognition using Centre Symmetric Local Binary Pattern Descriptors", 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1- 6.

[20].Liu, C., Ying, J., Yang, H. et al. "Improved human action recognition approach based on two-stream convolutional neural network model". pp. 1327–1341(2021).

[21].C. J. Dhamsania and T. V. Ratanpara, "A surveyon Human action recognition from videos", 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 2016, pp.1-5.

[22].Bhorge, Sidharth & Bedase, Deepak. (2018). "Multi View Human Action Recognition Using" HODD: Second International Conference, ICACDS 2018, Dehradun, India, April 20-21, 2018.

[23].Velastin, Sergio & Murtaza, Fiza & Yousaf, Muhammad Haroon. (2016)." Multi-view Human Action Recognition using 2D Motion Templates based on MHIs and their HOG Description".

[24].W. Wu and J. Yu, "An Improved Deep Relation Network for Action Recognition in Still Images", ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 2450- 2454.

[25].Lavinia, Yukhe & Vo, Holly & Verma, Abhishek. (2020). "New colour fusion deep learning model for large-scale action recognition". International Journal of Computational Vision and Robotics.

[26].Neziha Jaouedi, Noureddine Boujnah, Med Salim Bouhlel, "A new hybrid deep learning model for human action recognition" , Journal of King Saud University – Computer and Information Sciences, Volume 32, 2020, pp. 447-453.

[27].A.B. Sargano, X. Wang, P. Angelov and Z. Habib, "Human action recognition using transfer learning with deep representations", 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp.463-469.

[28].Saima Nazir, Muhammad Haroon Yousaf,Jean-Christophe Nebel, Sergio A. Velastin, "A Bag of Expression framework for improved human action recognition, Pattern Recognition Letters", Volume 103, 2018, pp. 39-45.

[29].A. B. Sargano, X. Wang, P. Angelov and Z. Habib, "Human action recognition using transfer learning with deep representations", 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp.463-469.

[30].C. Huang, C. Wang and J. Wang, "Human action recognition system for elderly and children care using three stream ConvNet", 2015 International Conference on Orange Technologies (ICOT), 2015, pp. 5-9.

[31].K. -P. Chou et al., "Robust Feature-Based Automated Multi-View Human Action Recognition System", in IEEE Access, vol. 6, pp. 15283-15296, 2018.

[32].A. Ta, C. Wolf, G. Lavoué, A. Baskurt and J. Jolion, "Pairwise Features for Human Action Recognition", 2010 20th International Conference on Pattern Recognition, 2010, pp. 3224-3227.

Cite this Article

Pavan M, Deepika D, Divyashree R, Kavana K, Pooja V Biligi, "Survey On Feature Extraction Approach for Human Action Recognition in Still Images and Videos", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 3, pp. 359-369, May-June 2022. Available at doi : https://doi.org/10.32628/CSEIT228392
Journal URL : https://ijsrcseit.com/CSEIT228392