ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/247-252
Pilli Vijay Babu / International Journal of Management Research & Review

Human Action Recognition STIP Using NB, KNN and LSTM, CNN

PILLI VIJAY BABU
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
CH.JEEVAN BABU
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract ***This paper compares three practical, reliable, and generic systems for multi view video-based human action recognition, namely, the NB, KNN and LSTM, CNN. To describe the different actions performed in different views, view-invariant features are proposed to address multi view action recognition. These features are obtained by extracting the holistic features from different temporal scales which are modelled as points of interest which represent the global spatial-temporal distribution. Experiments and cross-data testing are conducted on the datasets. The experiment results show that the proposed approach outperforms the existing methods on the datasets. INDEX TERMS: Multi-view video, action recognition, feature extraction, background subtraction, classification, machine learning.***

I. INTRODUCTION

Recently, human action recognition research has brought many challenges in the areas of sports, security and personal health care systems. Automatic video analysis systems which can recognize events related to human actions are becoming necessary in different industry areas. Therefore, human action recognition has become a hot research area in computer vision and there have been many papers published on this and many real-world applications have been developed, such as searching for the structure of large video archives, gesture recognition, video indexing, and video surveillance.

Human-computer interaction, in particular, is a crucial application in action recognition research. Visual cues are a significant part of human computer interaction to enable better communication between humans and computers; hence researchers utilize visual cues to recognize gestures and actions. Most of the recent action recognition work samples an action sequence manually before it can be recognized in a film. However, it is not practical to manually set the beginning and ending of an action sequence of the film previously. Therefore, a practical recognition system needs to be able to automatically separate many actions in an image sequence.

The current published methods for action recognition often sample an action sequence manually before it is recognized in a film [8]–[10]. However, it is not practical that setting the beginning and end of an action sequence of the film previously. Therefore, a practical recognition system needs to separate many actions at an image sequences automatically. Moreover, actions can be performed as different subjects such as size, posture, motion and clothing, which is still a challenging problem for several reasons, such as illumination, occlusion, shadow, camera movement or other environment changes. In addition, the actions depend on or involve objects which could add another layer of variability. As a consequence, action recognition methods often assume that the action is captured under restricted and simplified environments such as static backgrounds, non-complicated action classes and static cameras.

In particular, frequently moving the camera to an unknown position is the main cause of view variations. Similar to observing static objects from multi-view points, the actions may appear to be different from different angles. On the other hand, a moving camera could also affect the action appearance by incorporating dynamic view changes. Therefore, an action recognition system should be robust against environment and view-point changes when capturing an action sequence.

Fig.1.Proposed human-action recognition module.

II. LITERATURE SURVEY

[1] D. Geronimo, and H. Kjellstrom, "Unsupervised Surveillance Video Retrieval Based on Human Action and Appearance," International Conference on Pattern Recognition, pp.4630, 4635, 24-28, August 2014.

Forensic video analysis is the offline analysis of video aimed at understanding what happened in a scene in the past. Two of its key tasks are the recognition of specific actions, e.g., walking or fighting, and the search for specific persons, also referred to as re-identification. Although these tasks have traditionally been performed manually in forensic investigations, the current growing number of cameras and recorded video leads to the need for automated analysis. In this paper we propose an unsupervised retrieval system for surveillance videos based on human action and appearance. Given a query window, the system retrieves people performing the same action as the one in the query, the same person performing any action, or the same person performing the same action. We use an adaptive search algorithm that focuses the analysis on relevant frames based on the inter-frame difference of foreground masks. Then, for each analyzed frame, a pedestrian detector is used to extract windows containing each pedestrian in the scene. For each detection, we use optical flow features to represent its action and color features to represent its appearance. These extracted features are used to compute the probability that the detection matches the query according to the specified criterion. The algorithm is fully unsupervised, i.e., no training or constraints on the appearance, actions or number of actions that will appear in the test video are made. The proposed algorithm is tested on a surveillance video with different people performing different actions, providing satisfactory retrieval performance.

[2] K.T. Song, and W.J. Chen, "Human activity recognition using a mobile camera," International Conference on Ubiquitous Robots and Ambient Intelligence, pp.3,8, 23-26, November 2011.

This paper presents a vision-based human activity recognition system using a mobile camera. This system aims to enhance human-robot interaction in a home setting for applications such as health care and companion. In the first place, the camera needs to find a human in image frames. The body pose is classified for the detected human. Then the human activity is recognized by combining information of human pose, human location and elapsed time. In order to determine the situated place of the person in a home setting, a novel space-boundary detection method is proposed in this paper. This method uses features in the environment to automatically set space boundary in the image such that human location in the environment can be obtained. In the integrated experiments, human pose recognition rate of five poses(standing, walking, sitting, squatting, lying) is 94.8%. Experiments of human activity recognition in a home setting have been conducted to verify the performance of the proposed method by using a mobile camera from different view angles and positions in a home setting. The experimental results reveal that the space boundaries are detected as expected and satisfactory results are obtained.

[3] I. Everts, J.C. van Gemert, and T. Gevers, "Evaluation of Color Spatiotemporal Interest Points for Human Action Recognition," IEEE Transactions of Image Processing, vol.23, no.4, pp.1569,1580, April 2014.

This paper considers the recognition of realistic human actions in videos based on spatio-temporal interest points (STIPs). Existing STIP-based action recognition approaches operate on intensity representations of the image data. Because of this, these approaches are sensitive to disturbing photometric phenomena, such as shadows and highlights. In addition, valuable information is neglected by discarding chromaticity from the photometric representation. These issues are addressed by color STIPs. Color STIPs are multichannel reformulations of STIP detectors and descriptors, for which we consider a number of chromatic and invariant representations derived from the opponent color space. Color STIPs are shown to outperform their intensity-based counterparts on the challenging UCF sports, UCF11 and UCF50 action recognition benchmarks by more than 5% on average, where most of the gain is due to the multichannel descriptors. In addition, the results show that color STIPs are currently the single best low-level feature choice

for STIP-based approaches to human action recognition.We have reformulated STIP detectors and descriptors to incorporate multiple photometric channels in addition to image intensities, resulting in color STIPs.

[4] W. Heng, A. Klaser, C. Schmid, and L. Cheng-Lin, "Action recognition by dense trajectories," IEEE Conference on Computer Vision and Pattern Recognition, pp.3169,3176,20-25, June 2011.

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 This paper has introduced an approach to model videos by combining dense sampling with feature tracking. Our dense trajectories are more robust than previous video descriptions. They capture the motion information in the videos efficiently and show improved performance over state-of-the-art approaches for action classification. Wehave also introduced an efficient solution to remove camera motion by computing motion boundaries descriptors along the dense trajectories. This successfully segments the relevant motion from background motion, and outperforms previous video stabilization methods. Our descriptors combine trajectory shape, appearance, and motion information. Such a representation has shown to be efficient for action classification, but could also be used in other areas, such as action localization and video retrieval.

III. EXISTING METHOD

For the human action recognition support vector machine is used in existing work. Wei-Ta Chu and Shang-Yin Tsai [1] proposed a framework to extract rhythm information in dance videos and music, and accordingly correlate them based on rhythmic representation. In this work, the investigation of how rhythm information can be found and utilized in street dance videos. From the visual track, periodic motion changes of dancer's movement are extracted, which constitute "rhythm of motion" (ROM). From music, rhythm is constructed based on periodic properties of music beats. From dancer's movement, they constructed motion trajectories, detect turnings and stops of trajectories, and then estimate rhythm of motion (ROM). For music, beats are detected to describe rhythm of music. tempo. In dance videos, ROM is a clue about how a dancer interprets a music piece. Dancers usually divide the music into segments of "eight beats", and then design dancing steps for each segment. Although different dancers have varied styles on poses or body movement, they make emphasized stop or turning at boundaries of eight-beat segments. To extract motion trajectories, considered motion on feature points rather than all pixels in video frames. They adopted the Shi-Tomasi (ST) corner detector, because it is shown to be robust under affine transformation and can be implemented easily. And applied the Pyramid Lucas-Kanade (PLK) optical flow detection method to predict motion in various scales.

Andargie Mekonnen, Dr. Subramanian Karpaga Selvi, DrVajeravel Sam path Kumar, Bisratie Tesfaye [2] proposed optical flow method on complex motion. Optical flows are extracted in Ethiopian traditional dance video using Lucas-Kanade and Horn-Schunck methods. The optical flows estimated from these methods have three dimensions which is complex for real-time application. The magnitude of optical flow only considered to reduce complexity in training and to

make it suitable for real time content based video classification. The effectiveness of the method is evaluated by artificial neural network (ANN) to classify videos with Ethiopian traditional dances. In the study the Lucas Kanade method outperformed by scoring 82.0% of total accuracy whereas Horn-Schunck method and the combination of Lucas-Kanade and Horn Schunck methods less performed by registering 74.6% and 52.8% of overall accuracy respectively.

PROPOSED METHOD

For the proposed methodology for the human action recognition we used CNN.

4.1 DATASETS

Group activity detection aims to solve the problem of classifying group activities into certain categories. In our work, we aims at group dance activity recognition. In this group dances we mainly focus on tollywood group dances. Till now we do not have any benchmark datasets or databases. So, in our work we collected few tollywood group dances video data.

4.2 FEATURE EXTRACTION

The characteristics of the objects such as shape, silhouette, colours and motions are extracted and represented in some form of features. Generally speaking, the features can be categorized as four groups, space-time information, frequency transform, local descriptors and body modeling.

Figure 2. The categories for feature extraction and representation.

There are two popular approaches for feature extraction in group activity detection. One is using whole video sequences (i.e., global features). The other is analyzing certain regions of video sequences (i.e., local features).The space-time volume (STV) is built as the image features by concatenating the consecutive silhouette of objects along the time axis. The extracted 3D XYT volume (along x-y spatial coordinates and time) can capture the continuity of human action. But the STV is limited on non-periodic activities. The discrete Fourier transforms (DFT), which has been widely used to represent information about the geometric structure of the object.

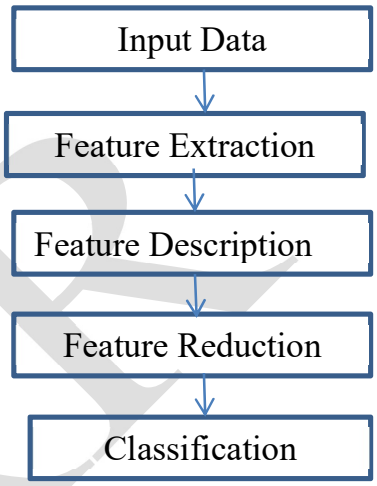
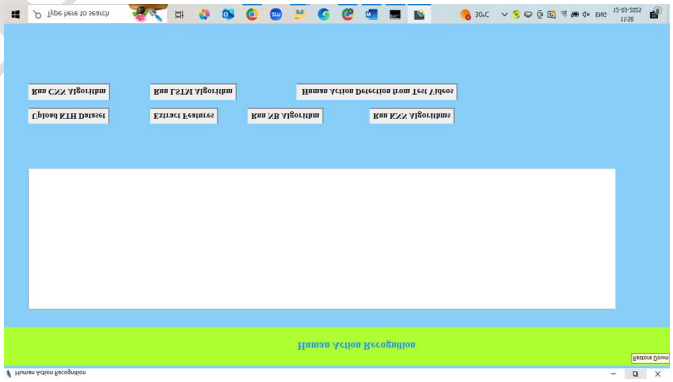
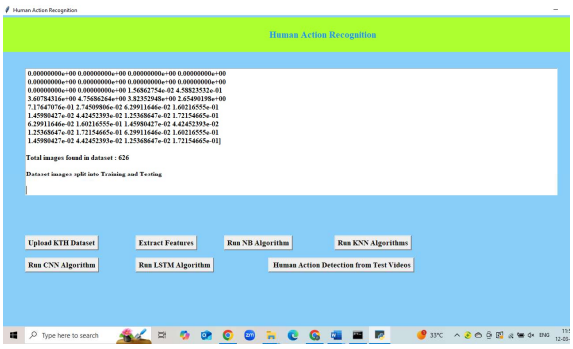


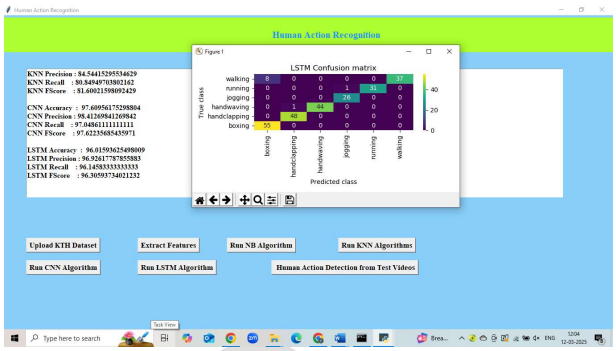
Fig.5 Flowchart

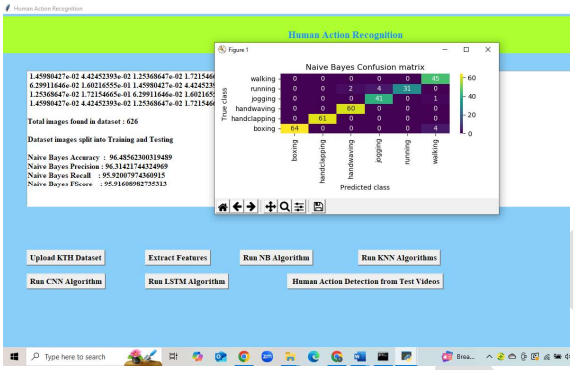
IV. RESULT

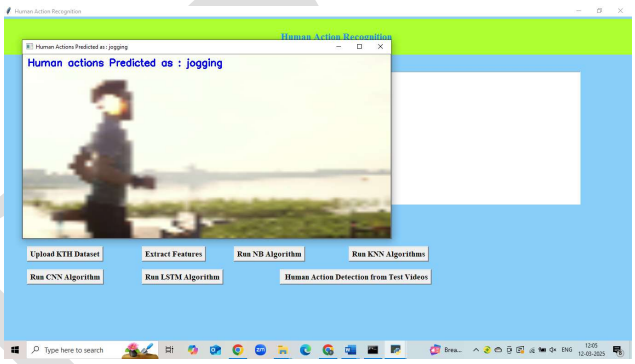


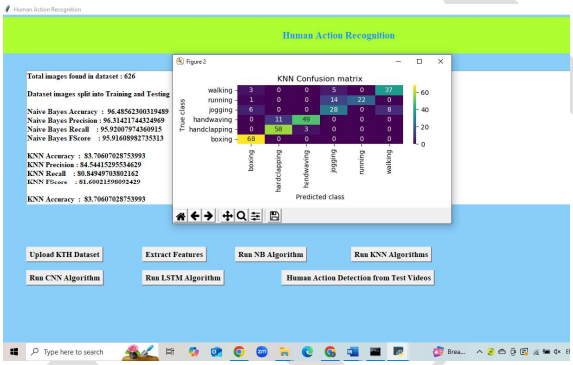
ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/247-252
Pilli Vijay Babu / International Journal of Management Research & Review

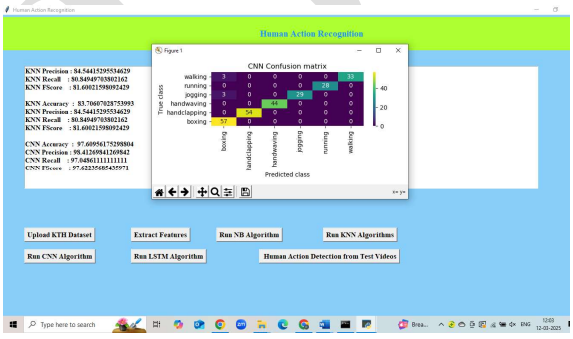












V. CONCLUSION

This paper presents an approach for real-world applications which automatically labels the beginning and ending of an action sequence. The system uses the proposed view-invariant features to address multi-view action recognition from different perspectives for accurate and robust action recognition. The view-invariant features are obtained by extracting holistic features from different temporal scale clouds, which are modeled on the explicit global, spatial and temporal distribution of interest points. The experiments datasets demonstrate that using viewinvariant features obtained by extracting holistic features from clouds of interest points is highly discriminative and more robust for recognizing actions under different view changes. The experiments also show the proposed approach performs well with cross-tested datasets using previously trained data, which means there is no need to re-train the system if the scenario changes.

REFERENCES

[1] K. G. Derpanis, M. Sizintsev, K. J. Cannons, and R. P. Wildes, ‘‘Action spotting and recognition based on a spatiotemporal orientation analysis,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 527–540, Mar. 2013.

[2] A. Gilbert, J. Illingworth, and R. Bowden, ''Action recognition using mined hierarchical compound features,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 883–897, May 2011.

[3] L. Liu, L. Shao, X. Zhen, and X. Li, ''Learning discriminative key poses for action recognition,'' IEEE Trans. Cybern., vol. 43, no. 6, pp. 1860–1870, Dec. 2013.

[4] Y. Yang, I. Saleemi, and M. Shah, ''Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 7, pp. 1635–1648, Jul. 2013.

[5] Z. Jiang, Z. Lin, and L. S. Davis, ''Recognizing human actions by learning and matching shape-motion prototype trees,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 533–547, Mar. 2012.

[6] K. Guo, P. Ishwar, and J. Konrad, ''Action recognition from video using feature covariance matrices,'' IEEE Trans. Image Process., vol. 22, no. 6, pp. 2479–2494, Jun. 2013.

[7] Y. Chen, Z. Li, X. Guo, Y. Zhao, and A. Cai, ''A spatio-temporal interest point detector based on vorticity for action recognition,'' in Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW), Jul. 2013, pp. 1–6.

[8] S. Samanta and B. Chanda, ''Space-time facet model for human activity classification,'' IEEE Trans. Multimedia, vol. 16, no. 6, pp. 1525–1535, Oct. 2014.

[9] Z. Moghaddam and M. Piccardi, ''Histogram-based training initialisation of hidden Markov models for human action recognition,'' in Proc. 17th IEEE Int. Nat. Conf. Adv. Video Signal Based Surveill. (AVSS), Sep. 2010, pp. 256–261.

[10] Y. Wang, L. Wu, and X. Huang, ''Action recognition using tri-view constraints,'' in Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS), Aug. 2011, pp. 107–112.

[11] M. N. Kumar and D. Madhavi, ''Improved discriminative model for viewinvariant human action recognition,'' Int. J. Comput. Sci. Eng. Technol., vol. 4, no. 3, pp. 1263–1270, 2013.

[12] F. Zhang, Y. Wang, and Z. Zhang, ''View-invariant action recognition in surveillance videos,'' in Proc. 1st Asian Conf. Pattern Recognit. (ACPR), Nov. 2011, pp. 580–583.

[13] T. Guha and R. K. Ward, ''Learning sparse representations for human action recognition,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 8, pp. 1576–1588, Aug. 2012.

[14] N. Ikizler-Cinbis and S. Sclaroff, ''Web-based classifiers for human action recognition,'' IEEE Trans. Multimedia, vol. 14, no. 4, pp. 1031–1045, Aug. 2012.

[15] D. Wu and L. Shao, ''Silhouette analysis-based action recognition via exploiting human poses,'' IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 2, pp. 236–243, Feb. 2013.

[16] C. Schuldt, I. Laptev, and B. Caputo, ''Recognizing human actions: A local SVM approach,'' in Proc. 17th Int. Conf. Pattern Recognit., vol. 3. Aug. 2004, pp. 32–36.

[17] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, ''Actions as space-time shapes,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 12, pp. 2247–2253, Dec. 2007.

[18] S. Singh, S. A. Velastin, and H. Ragheb, ''MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods,'' in Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Sep. 2010, pp. 48–55.

[19] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, ''Recognizing action at a distance,'' in Proc. IEEE, Oct. 2003, p. 726.

[20] A. Fathi and G. Mori, ''Action recognition by learning mid-level motion features,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2008, pp. 1–8.

[21] C. Rao and M. Shah, ''View-invariance in action recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2. Jun. 2001, p. II.

[22] A. Ali and J. K. Aggarwal, ''Segmentation and recognition of continuous human activity,'' in Proc. IEEE Workshop IEEE Detection Recognit. Events Video, Jul. 2001, pp. 28–35.

[23] D. Ramanan and D. A. Forsyth, ''Automatic annotation of everyday movements,'' in Proc. Adv. Neural Inf. Process. Syst., 2004, pp. 1547–1554.

[24] Y. Sheikh, M. Sheikh, and M. Shah, ''Exploring the space of a human action,'' in Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV), vol. 1. Oct. 2005, pp. 144–149.

[25] Y. Ke, R. Sukthankar, and M. Hebert, ''Efficient visual event detection using, volumetric features,'' in Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV), vol. 1. Sep. 2005, pp. 166–173.