**LITERATURE SURVEY**

**P. Pareek and A. Thakkar, “A survey on video-based Human Action**

**Recognition: recent updates, datasets, challenges, and applications,”**

**Artif Intell Rev, vol. 54, no. 3, pp. 2259–2322, Mar. 2021, doi:**

**10.1007/s10462-020-09904-8.**

Human Action Recognition (HAR) involves human activity monitoring task in different areas of medical, education, entertainment, visual surveillance, video retrieval, as well as abnormal activity identification, to name a few. Due to an increase in the usage of cameras, automated systems are in demand for the classification of such activities using computationally intelligent techniques such as Machine Learning (ML) and Deep Learning (DL). In this survey, we have 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 namely, content-based video summarization, human–computer interaction, education, healthcare, video surveillance, abnormal activity detection, sports, and entertainment. The advantages and disadvantages of action representation, dimensionality reduction, and action analysis methods are also provided. The paper discusses challenges and future directions for HAR.

**P. K. Singh, S. Kundu, T. Adhikary, R. Sarkar, and D. Bhattacharjee,**

**“Progress of Human Action Recognition Research in the Last Ten**

**Years: A Comprehensive Survey,” Archives of Computational**

**Methods in Engineering, vol. 29, no. 4, pp. 2309–2349, Jun. 2022,**

**doi: 10.1007/s11831-021-09681-9.**

Human Action Recognition (HAR) has achieved a remarkable milestone in the field of computer vision. Apart from its varied applications in human–computer interactions, surveillance systems and robotics, in recent times, it has extended its applicability in the fields like healthcare, multimedia retrieval, social networking, and education as well. Over the years, various approaches have been proposed by researchers to develop systems for HAR. In this context, this survey mainly deals with the various categories of approaches that have been proposed for HAR in the last ten years. To be specific, HAR techniques range from conventional machine learning methods to recently popular deep learning methods, and this field is growing fast. Hence, there is a need for frequent surveys to keep track of the latest techniques employed, the corresponding results achieved, and the problems that remain to help show the path forward, which this survey aims to accomplish. HAR can be classified into two divisions: unimodal and multimodal depending on the type of input vectors—unimodal implies that the data comes from a single source, while multimodal means the input dataset is from more than one source. This survey covers significant methods developed for the unimodal HAR in the past decade. The unimodal methods have been classified and described using the concepts of machine learning. Further, numerous models suggested for HAR using deep learning have been discussed elaborately. A list of different feature extractors and a detailed account of some majorly used video and still-image datasets have also been described in this survey, along with some useful insights into future work scope.

**A. Ladjailia, I. Bouchrika, H. F. Merouani, N. Harrati, and Z.**

**Mahfouf, “Human activity recognition via optical flow: decomposing**

**activities into basic actions,” Neural Comput Appl, vol. 32, no. 21,**

**pp. 16387–16400, Nov. 2020, doi: 10.1007/s00521-018-3951-x**

Recognizing human activities using automated methods has emerged recently as a pivotal research theme for security-related applications. In this research paper, an optical flow descriptor is proposed for the recognition of human actions by considering only features derived from the motion. The signature for the human action is composed as a histogram containing kinematic features which include the local and global traits. Experimental results performed on the Weizmann and UCF101 databases confirmed the potentials of the proposed approach with attained classification rates of 98.76% and 70%, respectively, to distinguish between different human actions. For comparative and performance analysis, different types of classifiers including Knn, decision tree, SVM and deep learning are applied to the proposed descriptors. Further analysis is performed to assess the proposed descriptors under different resolutions and frame rates. The obtained results are in alignment with the early psychological studies reporting that human motion is adequate for the perception of human activities.

**K. Simonyan and A. Zisserman, “Two-Stream Convolutional**

**Networks for Action Recognition in Videos.”**

We investigate architectures of discriminatively trained deep Convolutional Networks (ConvNets) for action recognition in video. The challenge is to capture the complementary information on appearance from still frames and motion between frames. We also aim to generalise the best performing hand-crafted features within a data-driven learning framework. Our contribution is three-fold. First, we propose a two-stream ConvNet architecture which incorporates spatial and temporal networks. Second, we demonstrate that a ConvNet trained on multi-frame dense optical flow is able to achieve very good performance in spite of limited training data. Finally, we show that multitask learning, applied to two different action classification datasets, can be used to increase the amount of training data and improve the performance on both. Our architecture is trained and evaluated on the standard video actions benchmarks of UCF-101 and HMDB-51, where it is competitive with the state of the art. It also exceeds by a large margin previous attempts to use deep nets for video classification