World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:10, 2024

3D Human Reconstruction over Cloud Based Image Data via AI and Machine Learning

Kaushik Sathupadi, Sandesh Achar

***Abstract*—**Human action recognition (HAR) modeling is a critical task in machine learning. These systems require better techniques for recognizing body parts and selecting optimal features based on vision sensors to identify complex action patterns efficiently. Still, there is a considerable gap and challenges between images and videos, such as brightness, motion variation, and random clutters. This paper proposes a robust approach for classifying human actions over cloud-based image data. First, we apply pre-processing and detection, human and outer shape detection techniques. Next, we extract valuable information in terms of cues. We extract two distinct features: fuzzy local binary patterns and sequence representation. Then, we applied a greedy, randomized adaptive search procedure for data optimization and dimension reduction, and for classification, we used a random forest. We tested our model on two benchmark datasets, AAMAZ and the KTH Multi-view Football datasets. Our HAR framework significantly outperforms the other state-of-the-art approaches and achieves a better recognition rate of 91% and 89.6% over the AAMAZ and KTH Multi-view Football datasets, respectively.

***Keywords*—**Computer vision, human motion analysis, random forest, machine learning.

I. Introduction

HUMAN motion analysis characterizes a significant promising area of research within a video processing and AI, with the potential to transform different aspects of human life. Additionally, it makes possible for machines to mimic how a human behaves in an environment and, as a result, to take the initiative in any circumstances across various data sources, including cloud-based RGB data, real-time data, IoT-based video, and sensor data [1], [2]. This capability to replicate human intelligence in decision making strategies and reshape to new scenarios is a major take in the advancement of autonomous systems. Existing applications such as semantic video indexing, human action recognition, and motion detection can automate intelligent systems entirely or partially over cloud data. In addition, these programs demand a localization method and action detection framework for the identifying competence technology to perform more effectively [3], [4].

Digitalizing the video processing-based innovative system is one of the significant functions of cloud data and crowded environments. The advanced monitoring system allows streaming video material to utilize the same channel without interference from other data. Due to the development of connectivity, archiving, and image compression, the computer program has iteratively enhanced its technical progress. Utilizing an intelligent computerized monitoring system in a hostile environment enhances law enforcements' capacity to take appropriate action in various situations [5]. Identifying and capturing visual patterns from the human face and skeleton is a significant difficulty in image compression and video monitoring. Detecting the individual requires the following steps: monitoring, identification, categorization, and authentication of body motion [6], [7]. The intelligent system uses computer vision and information processing to confirm the person's authenticity under cloud technology-based video data [8]-[10].

Various applications and systems have been proposed to address the different remaining challenges for HMR-based intelligent systems. Most systems were based on traditional feature extraction methods such as optical flow and distance features. Additionally, they adopted standard machine learning algorithms such as SVM and decision trees. These systems have achieved a low accuracy rate due to the high volume of features and feature vector dimensions. In some way, various systems provide a high error rate of human detection and human action recognition due to weak classification method selection; different systems face complex, crowded data, and occlusion issues in different data sets and frame sequences.

In this paper, we provide a functional approach to address these types of issues. We consider cloud-based data containing crowded data of humans in various life log scenes. First, we performed video-to-frame transformation and resized the image; with the help of k-means clustering, we extracted complete information, such as a human in video scenes. After that, the template-based method is applied to detect humans and spread the hierarchal skeleton model to analyze human motion, and the next step is the extraction of features. Spatial and binary features are the primary concerns. Finally, we applied a greedy, randomized adaptive search procedure; we used a random forest for abnormal action classification and identification. The main contributions of this paper are as follows.

- Our proposed system involves cloud-based and real-time video data from two benchmark datasets across various intelligent systems. Additionally, these two benchmark datasets - AAMAZ and the KTH Multi-view football datasets- were subjected to comprehensively analyzed for a human pose estimation system. The results of the experiments attain a better recognition rate.
- Features extraction and hierarchical skeleton system accurately understood human posture and actions.

Sandesh Achar is Senior Manager and Software Engineer at Walmart Global Tech, Sunnyvale, California, 94086, USA (e-mail: sandeshachar26@gmail.com).

- Deep data optimization through GRASP and random forests are used for classification to get optimized results.

The remaining sections are organized as follows: Section II includes a collection of previous human-based recognition techniques. In Section III, we discussed the method and framework for activity recognition. Section IV provides the results and experimental setup. Finally, in Section V, conclusions and research directions are presented.

II. Related Work

Numerous studies have already been conducted on human action recognition (HAR). This section introduces some literature work on HAR and cloud-based distributed systems for evaluating multimedia data, which are relevant to the proposed method. The essential element of HAR is acquiring the time sequence of video shorts. In this context, Gaidon et al. [11] proposed an action sequence framework that presents an action as a series of histograms of visual descriptors. It is distinctive, unlike the bag-of-features; it presents an action with multiple components rather than an individual mean representation. The experimental results revealed that the method significantly outperformed the conventional bag-of-features and other state-of-the-art methods. Wang et al. [12] presented an LHM model for identifying challenging actions. The primary purpose of this method is to calculate the variables of LHM. In addition, they demonstrated a cascade inference approach to enhance the effectiveness of activity classification. During the inference process, each action's beginning and starting times are represented by these variables and automatically identified. The experimental results have shown that the proposed technique incorporated with dense features attained better recognition performance on multiple benchmark datasets. In another study, Zhang et al. [13] introduced a novel cloud-based framework that can provide a stable solution for intelligent analysis and storage of image sequences, referred to as BiF architecture. The novel system can efficiently manage continuous surveillance video-based data by integrating real-time assessment, stream processing, and distributed storage which are flawlessly fused to fulfill the processing needs of video data. The results determine that the proposed system is robust regarding speed, space, and fault tolerance. Lan et al. [14] proposed a hierarchical system of mid-level action elements (MAE) for action recognition and parsing in image sequences. First, the proposed system analyzes input source video data into a tree structure based on MAEs at numerous scales. Each MAE identifies a human action-related spatiotemporal segment in the scene. In this context, human actions and labels at diverse granularity levels are finer-grained and collectively inferred. The results revealed significant performance over standard techniques and numerous publicly accessible datasets.

Azhar [15] demonstrated a state-of-the-art method for recognizing human actions. Additionally, the author analyzed Apache Spark to tackle the challenges of HAR. Furthermore, the author created a unique feature descriptor, an adaptive local motion-based descriptor capable of extracting texture and movement features and generating persistent patterns. Chou et al. [16] proposed a technique for labeling the initial and ending points of a human activity sequence in a real-world environment. Additionally, the proposed system employed the view-invariant descriptors to solve multi-view-based HAR from diverse perspectives. Furthermore, the view-invariant-based descriptors are generated by abstracting holistic attributes from various time-based scale clouds predicated on the explicit global, geographical, and time series of interest points. The results on benchmark datasets revealed that incorporating view-invariant features acquired by abstracting holistic attributes from these points is significantly distinct and more effective for recognizing human actions under different perspectives.

III. System Design

In this section, we elaborate on the complete image of our proposed HMR system. We begin with the pre-processing and data normalization phases, followed by the human detection, skeleton, and feature extraction segments of the proposed method. The next step involves data optimization using GRASP, culminating in classification via the random forest, Fig. 1 illustrates the workflow and layout of the proposed system.

A. Pre-processing

In this phase, we discuss the pre-processing techniques over cloud-based video data; video-to-frame conversion is the initial step, and we resize the extracted frames to avoid extra computational cost and power. Then, the source image is routed to the object verification stage, where the Viola-Jones technique extracts the objects [15]. The Viola-Jones approach is predominantly employed in detection and object recognition, characterized by a slower overall process while the detection step itself is rapid. Finally, Viola-Jones uses the transformation function to retrieve the rectangular-shaped Haar-like characteristics as follows:

$$F(x) = \alpha 1F1(x) + \alpha 2F(x) \quad (1)$$

Detection and recognition are conducted within the sliding window, where both maximum and minimal dimensions of the original image are specified, and a feature set is constructed for each scale factor. Each filter includes a descriptor that is applied to aggregate the characteristics. The robust learners are concatenated, and the filter is then utilized to determine the items. The result of this identification process, denoted as $o1$, indicates the presence of an object, with a frequency of either 0 or 1. Following this, we apply k-means for optimized background subtraction and human shape extraction. The k-means method is an unsupervised methodology that is utilized to sequence the preferences and interests of the environment. It clusters or divides the gathered information into *K*-clusters or segments depending on the *K*-centroids. The automated system is used if there are unmarked data (i.e., statistics without characterized groups or categories). The objective is to discover a certain degree of familiarity in the statistical information with the group followed by K. The number of squared values across all endpoints and the support vectors must be kept to a minimum when using k-means segmentation.

$$J = \sum_{r=1}^{c} \sum_{i=1}^{n} ||x_i^j - c_r||^2 \quad (2)$$

where r = objective function, c = number of clusters, n = number of cases, x case, and c = centroid. Fig. 2 shows the detailed results.

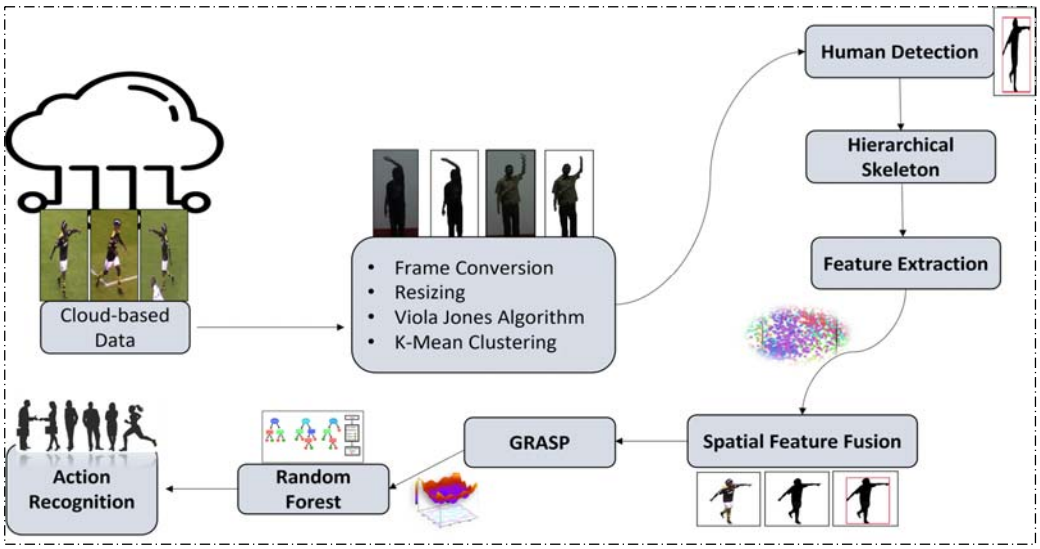


Fig. 1 The architectural layout of the proposed HMR framework

Fig. 2 The optimized results of pre-processing and background subtraction over the AAMAZ dataset

B. Human Detection and Hierarchical Skeleton

Connected components are a straightforward and typical technique to illustrate the body. For instance, a descending extremity is attached to an upper extremity, which is then attached to the torso. These designs are frequently employed in digital effects. This paradigm, therefore, is centered on "part-of" interactions. For instance, a lower limb and an upper limb are components of a body, which is a component of a limb. Whereas, the representation of a group of individuals consists of a hierarchical of eigenspaces in which an "increased" perception encompasses the critical features of a "lower-level" orthogonal projection as:

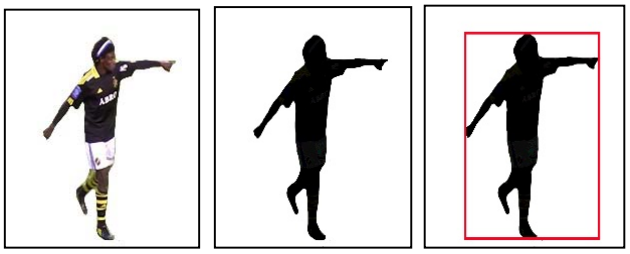
$$(p;\ U; A\ ; MK) \quad (3)$$

where p is the mean value of U and A is the engine value, and MK is the upper and lower limit. The best positions for a group of people doing a specific action such as walking, lying, and sitting are repeating characteristics of human behavior referred to as main postures. Bones, joints, and muscle protein synthesis make up efficient human performance. As they construct the skeleton, the bones of the human being are interconnected by joints; muscle contractions cling to the skeleton and bridge the joints. Utilizing the joint as a pivot point, the skeletal system requires variation in the bone situation due to skeletal suction, which is displacement. The skeletal system is a dynamic organ during activity.

Consequently, skeletal muscle is the active component of a human set of steps, whereas joints and muscles are present in the structure. To improve readability, we substitute the letter *I* with the cooperative notation. For instance, we can refer to the right hand as:

$$P_{RH}^t = (x_{RH}^t, y_{RH}^t, z_{RH}^t). \quad (4)$$

Fig. 3 shows the results of human detection.



(a) (b) (c)

Fig. 3 The human detection results (a) extracted background, (b) binary conversion, and (c) human detection over the KTH Multi-view Football dataset

C. Features Extraction

In this phase, we discuss the feature extraction method. We extract two state-of-the-art features: fuzzy local binary (LBP) patterns and sequence representation over the given dataset.

1. Features Extraction: Fuzzy Local Binary Patterns

FLBP combines LBP and fuzzy logic to strengthen the performance of discrimination and eliminate noise impacts. The crisp variant of LBP leverages the pixel attributes of neighboring pixels to understand each pixel. Using illumination adjustments makes altering the grayscale more durable, economical, and straightforward. The LBP structures efficiently represent the tiny texture features. LBP uses a hard segmentation algorithm to produce the pattern, has minimal classification performance, and is noise sensitive. The FLBP classification is an improved variant of LBP with highly distinguishable characteristics. In FLBP, each pixel represents any quantity of wavelet coefficients, which builds to the FLBP bin entropy. The affiliation function is developed as follows:

$$h_o(b) = \begin{cases} 0 \\ \frac{E - g_d + g_c}{2.E} \\ 1 \end{cases} \quad (5)$$

$$g_d \geq g_c + E$$

$$g_c < -E < g_d < g_c + E$$

$$g_d \leq -T$$

$$h_1(b) = 1 - h_o(\mathrm{b}) \quad (6)$$

The letters GD represent the neighboring pixel, $h1$ () and $h0$ () represent the hyperparameters, E represents the specified threshold used to determine the degree of fuzzy sets, and the distribution of LBP codes within the FLBP spectrum is determined by the linear model.

2. Spatial Features Extraction: Sequence Representation

The individual shapes are extracted for distinct motion sequences using geometrical post-processing, shade reduction, normalization, quantization, and feature extraction. Gait Energy Imaging (GEI) is thought to show a series of movements in an image that looks like:

$$A(a,b) = 1/D \sum_{f=1}^{b} C(a,b,f) \quad (7)$$

where D is the number of images in the full rotation, and b are the spatial coordinates, and $C(a.b.f)$ is the binary outline of the fth frame. The results of the spatial relationships, denoted as $o2$, also have a frequency f either 0 or 1. Finally, we arrange the features vector and pass it through GRASP for data mining and optimization.

D.Data Mining: GRASP

A simple method always selects the option that appears to have the optimal value at the time. In the hopes of arriving at a globally ideal answer, it makes a locally effective decision. Evolutionary algorithms do not always produce the best results (like 0-1 knapsack), but occasionally they do (e.g., minimum spanning tree). The next ingredient to be added to the response is often chosen by the greedy algorithm randomly [17]. The currently partial list of candidates, known as the Restricted Candidate List, will have elements added through a statistical shortlisting process. Unlike the stationary technique, which assigns a grade to materials before starting construction, this unique capability modifies the equation that directs the greedy system based on the material selected for development. GRASP is outlined in Algorithm 1.

Algorithm 01: GRASP

```
Step$ DesignSolution()
Solution^ = ∅
for  h solution designing not finished i
CandidateList$ = MakeCandidateList()
s = SelectElement(CandidateList)
Solution$ = Solution$ ∪{s}
ChangeGreedyFunction();
return  Solution
```

E. Classification: Random Forest

The classification technique involves a collection of regressions and classifications during which each indication is constructed using an orientation characteristic selected randomly from base classifiers. Furthermore, every tree applies a provided accordingly to determine its most powerful category for a training process. The random forest capable of enforcing throughout this work generates a tree including randomized features or a mix of properties. The methodology constructs a training dataset by continually producing N patterns with replacements. N is the quantity of the specific validation dataset utilized at each verification credible. Any visualizations (pixels) are categorized by determining the classification with the highest remarkable grades from the forest's classification methods. Establishing a tree structure required clearance criteria and a cleaning procedure [18]. Various ways are available for identifying properties for quantitative decision trees, with the major among these procedures being to present a numeric effect on the property. The Bilateral Understanding Ratio requirement and the Determined Coefficient are among the statistical method's most frequently applied filter feature selection processes. The random forest categorization methodology comprises Geometric Analysis, which examines the contamination of an indication relating to the classifiers as a constraint for obtaining the features. Fig. 4 shows the conceptual model of a random forest.

IV. System Design

A. Experimental Evaluation

This part contains the experimental work of the proposed HMR system. This work utilized a hierarchical skeleton system to get a more precise understanding of human motion and action. The performance of the proposed HMR framework was evaluated on two public benchmark datasets, AAMAZ and KTH Multi-view Football datasets, respectively. The dataset KTH Multi-view Football [19] includes 5907 images. Every image has 14 labeled components and a 2D ground truth stance. In addition, there are three distinct participants. Fig. 5 shows the example images of the KTH Multi-view Football dataset.

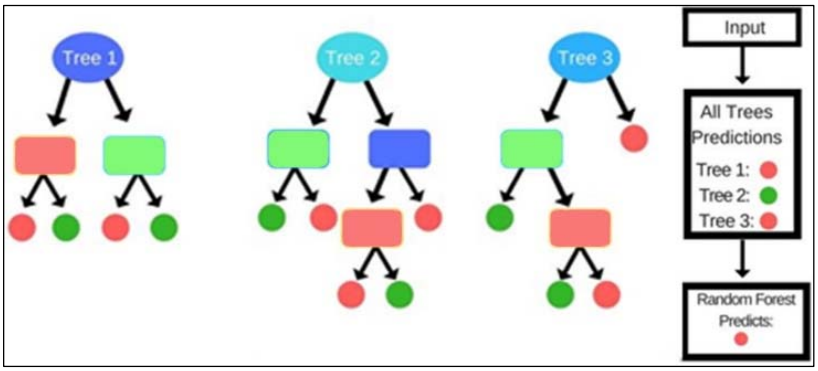


Fig. 4 The Conceptual Diagram of Random Forest

Fig. 5 The sample images of the KTH Multi-view Football dataset

The AAMAZ dataset is open to the public. It encompasses 11 kinds of activities, including walking, jogging, running, left-handed boxing, right-handed boxing, crawling, clapping, leaping, both hands waving, left-handed waving, and right-handed waving [20], [21]. Each activity is represented by approximately 14 video clips. Fig. 6 displays a selection of images from this dataset.

Fig. 6 The sample images of the AAMAZ dataset

Fig. 7 illustrates the confusion matrix for the AAMAZ dataset for 11 action classes with a 91% recognition rate. Table I represents the human body part recognition results of the KTH Multi-view Football dataset, which achieved a mean recognition rate of 89.16%. In Table I, we tested the HMR framework performance with two standard classifiers, genetic algorithm [22] and AdaBoost [23] techniques, via different performance metrics, including recall, precision, and F measures for all classes in the benchmark database. Table II estimates the human pose estimation over the KTH Multi-view Football dataset. Finally, Table III compares the proposed system performance over benchmark datasets with other state-of-the-art systems. After evaluating the results, the next step is to compare the performance of the HMR system with other state-of-the-art methods, as presented in Table II.

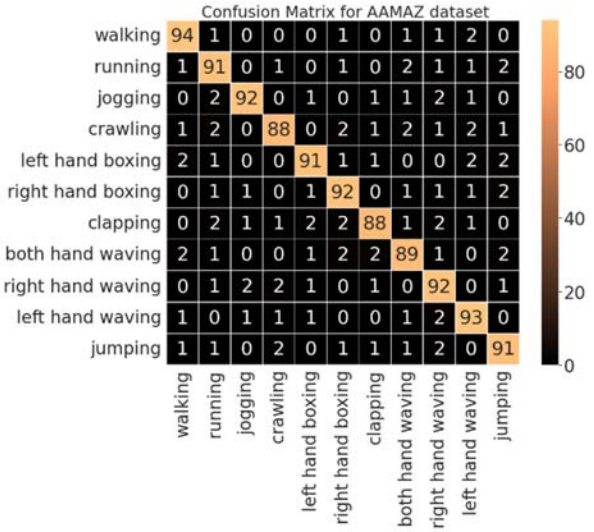


Fig. 7 Confusion Matrix of AAMAZ dataset over 11 actions

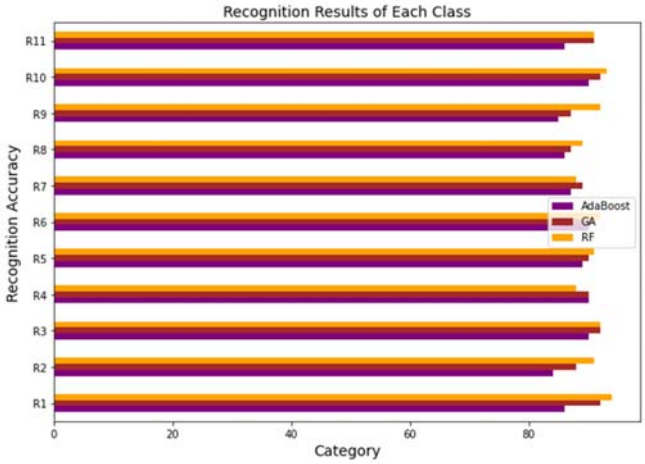


Fig. 8 Recognition comparison of the proposed HMR over two state-of-the-art classifiers over the AAMAZ dataset

TABLE I
COMPARISON OF THE EVALUATION METRICS (PRECISION, RECALL, AND F1 SCORE) OF THE HMR FRAMEWORK FROM THE AAMAZ DATABASE

Techniques	Random Forest			Genetic Algorithm			AdaBoost		
Gestures	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
G1	0.921	0.940	0.930	0.913	0.920	0.879	0.901	0.860	0.880
G2	0.883	0.910	0.896	0.872	0.880	0.864	0.864	0.840	0.852
G3	0.948	0.920	0.934	0.931	0.920	0.882	0.924	0.900	0.912
G4	0.926	0.880	0.902	0.915	0.900	0.857	0.899	0.894	0.896
G5	0.928	0.910	0.919	0.918	0.900	0.911	0.903	0.890	0.896
G6	0.92	0.920	0.920	0.914	0.910	0.880	0.902	0.900	0.901
G7	0.926	0.880	0.902	0.904	0.890	0.868	0.890	0.878	0.884
G8	0.890	0.890	0.890	0.885	0.870	0.867	0.870	0.860	0.865
G9	0.876	0.920	0.897	0.864	0.870	0.881	0.849	0.855	0.852
G10	0.930	0.930	0.930	0.912	0.920	0.877	0.902	0.908	0.905
G11	0.910	0.910	0.910	0.896	0.910	0.900	0.879	0.868	0.873
$W_{Avg.}$	0.914	0.910	0.911	0.902	0.899	0.900	0.889	0.877	0.803

TABLE II
HUMAN BODY PARTS RECOGNITION BASED PROPOSED SYSTEM OVER KTH MULTI-VIEW FOOTBALL DATASET

Human body parts	Total distance from ground truth	Recognition Accuracy (%)
left hand	9.8	90
right hand	10.2	92
left shoulder	11.9	84
right shoulder	11.6	87
left hip	10.8	89
right hip	11.7	91
upper head	9.1	89
torso	6.5	94
left foot	8.2	91
right foot	9.5	94
left knee	12.7	85
right knee	11.6	84
Mean recognition rate = 89.16%		

TABLE III
COMPARISON OF STATE-OF-THE-ART METHODS WITH THE PROPOSED SYSTEM

Methods	KTH Multi-view	Methods	AAMAZ
2D human pose estimation [24]	62.90%	Histograms of 3D joints [27]	71.10%
3DPS framework [25]	68.0%	Spatial-temporal attention [28]	78.33%
Human body parts detection [26]	84.01%	Spatial-temporal features [29]	81.18%
Proposed System	89.16%		91.0%

V. CONCLUSION

This paper presents a multi fused-based feature modeling for human motion reconstruction (HMR). The fuzzy LBP, sequence representation, and hierarchical skeleton are utilized for HMR effectively and robustly. Additionally, the multi-fused feature model is optimized through an adaptive research procedure, enhancing classification ability. Moreover, the proposed HMR system effectively learns the action pattern between successive frames. Furthermore, the experimental evaluation demonstrates that the proposed system outperforms other state-of-the-art methods to identify human actions on two challenging benchmark datasets: the KTH Multi-view Football and AAMAZ datasets. In the future, we will focus on HMR based on different views and numerous modalities and further improve the system.

REFERENCES

[1] Gutchess, D.; Trajkovics, M.; Cohen-Solal, E.; Lyons, D.; Jain, A.K. A background model initialization algorithm for video surveillance. In Proceedings of the Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; pp. 20017–142001.

[2] Mustafa, Z., Nsour, H., & ud din Tahir, S. B. (2023). Hand gesture recognition via deep data optimization and 3D reconstruction. *PeerJ Computer Science*, *9*, e1619.

[3] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang and J. Liu, "Human Action Recognition from Various Data Modalities: A Review," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022

[4] S. Tanberk, Z. H. Kilimci, D. B. Tükel, M. Uysal and S. Akyokuş, "A Hybrid Deep Model Using Deep Learning and Dense Optical Flow Approaches for Human Activity Recognition," *in IEEE Access*, vol. 8, pp. 19799-19809, 2020.

[5] M.-N. Chapel and T. Bouwmans, "Moving objects detection with a moving camera: A comprehensive review," *Comput. Sci. Rev.*, vol. 38, p. 100310, 2020.

[6] S. Angadi and S. Nandyal, "Human identification system based on spatial and temporal features in the video surveillance system," *Int. J. Ambient Comput. Intell.*, vol. 11, no. 3, pp. 1–21, 2020.

[7] Wieczorek, G.; Tahir, S.B.u.d.; Akhter, I.; Kurek, J. Vehicle Detection and Recognition Approach in Multi-Scale Traffic Monitoring System via Graph-Based Data Optimization. *Sensors* **2023**, *23*, 1731.

[8] Kumar, S., Shailu, A., Jain, A., & Moparthi, N. R. (2022). Enhanced method of object tracing using extended Kalman filter via binary search algorithm. *Journal of Information Technology Management*, *14*(Special Issue: Security and Resource Management challenges for Internet of Things), 180-199.

[9] Bhargavi, D.; Coyotl, E.P.; Gholami, S. Knock, knock. Who's there?--Identifying football player jersey numbers with synthetic data arXiv 2022, arXiv:2203.00734.

[10] Gholami, S.; Khashe, S. Alexa, Predict My Flight Delay. arXiv 2022, arXiv:2208.09921.

[11] Gaidon, A., Harchaoui, Z., & Schmid, C. (2013). Temporal localization of actions with actoms. *IEEE transactions on pattern analysis and machine intelligence*, *35*(11), 2782-2795.

[12] Wang, L., Qiao, Y., & Tang, X. (2013). Latent hierarchical model of temporal structure for complex activity classification. *IEEE Transactions on Image Processing*, *23*(2), 810-822.

[13] Zhang, W., Xu, L., Duan, P., Gong, W., Lu, Q., & Yang, S. (2015). A video cloud platform combing online and offline cloud computing technologies. *Personal and Ubiquitous Computing*, *19*, 1099-1110.

[14] Lan, T., Zhu, Y., Zamir, A. R., & Savarese, S. (2015). Action recognition by hierarchical mid-level action elements. In *Proceedings of the IEEE international conference on computer vision* (pp. 4552-4560).

[15] Azhar, S. (2021). *Automating Industrial Communication Standards selection by using a Knowledge-based systems* (Master's thesis).

[16] Chou, K. P., Prasad, M., Wu, D., Sharma, N., Li, D. L., Lin, Y. F., ... & Lin, C. T. (2018). Robust feature-based automated multi-view human action recognition system. *IEEE Access*, *6*, 15283-15296.

[17] Gupta, A., & Tiwari, R. (2015). Face detection using modified Viola jones algorithm. *International Journal of Recent Research in Mathematics Computer Science and Information Technology*, *1*(2), 59-66.

[18] Resende, M. G., & Ribeiro, C. C. (2010). Greedy randomized adaptive search procedures: Advances, hybridizations, and applications. *Handbook of metaheuristics*, 283-319.

[19] Itti, L., & Baldi, P. (2005, June). A principled approach to detecting surprising events in video. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 631-637). IEEE.

[20] Itti, L., & Baldi, P. (2005, June). A principled approach to detecting surprising events in video. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 631-637). IEEE.

[21] Nadeem, A., Jalal, A., & Kim, K. (2021). Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model. *Multimedia Tools and Applications*, *80*, 21465-21498.

[22] Jiang, Y.; Tong, G.; Yin, H.; Xiong, N. A Pedestrian Detection Method Based on Genetic Algorithm for Optimize XGBoost Training Parameters. *IEEE Access* 2019, *7*, 118310–118321.

[23] Subasi, A.; Dammas, DH; Alghamdi, RD; Makawi, R.A.; Albiety, E.A.; Brahimi, T.; Sarirete, A. Sensor Based Human Activity Recognition Using AdaBoost Ensemble Classifier. Procedia Comput. Sci. 2018, 140, 104–111.

[24] M. H. Oreaba, "Solving the confusion of body sides problem in two-dimensional human pose estimation", Master's Thesis, the American University, 2017.

[25] V. Belagiannis, S. Amin, M. Andriluka, B. Schiele, N. Navab and S. Ilic, "3D pictorial structures for multiple human pose estimation", *Proc. CVPR*, June 2014.

[26] H. W. Chen and M. MeGurr. "Moving human full body and body parts detection, tracking, and applications on human activity estimation." *SPIE Defense + Security*, 2016.

[27] Xia, L.; Chen, C.; Aggarwal, J.K. View invariant human action recognition using histograms of 3D joints. In Proceedings of the Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 20–27.

[28] Han, Y.; Chung, S.L.; Ambikapathi, A.; Chan, J.S.; Lin, W.Y.; Su, S.F. Robust human action recognition using global spatial-temporal attention for human skeleton data. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018.

[29] Das, S.; Chaudhary, A.; Bremond, F.; Thonnat, M. Where to focus on for human action recognition? In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 71–80.