

Unsupervised Human Activity Recognition Learning for Disassembly Tasks

MR. RUSHIKESH MURLIDHAR CHORMALE

Ashokrao Mane Group Of Institution, Vathar
Department of Computer Science and Engineering
rushikeshchormale001@gmail.com

MR. ONKAR SURESH KHARUDE

Ashokrao Mane Group Of Institution, Vathar
Department of Computer Science and Engineering
oskharude@gmail.com

MR. PRAVIN RAJENDRA KAMBLE

Ashokrao Mane Group Of Institution, Vathar
Department of Computer Science and Engineering
pk6618084@gmail.com

MR. NIKHIL ARJUN GHARPANKAR

Ashokrao Mane Group Of Institution, Vathar
Department of Computer Science and Engineering
nikhilag911@gmail.com

MR. SAHIL SANJAY KAGINKAR

Ashokrao Mane Group Of Institution, Vathar
Department of Computer Science and Engineering
sahilkaginkar1984@gmail.com

MR. ANIL B. DESAI

Assistant professor at Ashokrao Mane Group of
Institution, Vathar
Department of Computer Science and Engineering
abd@amgoi.edu.in

Abstract:

The rapid growth of electronic waste has increased the demand for efficient remanufacturing and recycling systems. In most remanufacturing industries, disassembly operations are still performed manually, resulting in low productivity, high operational cost and inconsistent performance. Human-robot collaboration has recently emerged as a practical solution for improving industrial disassembly processes. However, for robots to assist workers effectively, they must understand and recognize human activities accurately.

This paper presents an unsupervised human activity recognition framework for industrial disassembly tasks using video-based learning techniques. Unlike traditional supervised learning methods that require large amounts of manually labeled data, the proposed framework learns meaningful activity patterns directly from unlabeled video sequences. The system combines a Variational Autoencoder (VAE) for feature representation learning with a Hidden Markov Model (HMM) for temporal activity segmentation and classification.

The proposed framework is evaluated using a hard disk drive disassembly case study. Experimental analysis demonstrates that the framework achieves strong recognition performance while significantly reducing dependency on manual annotations. The proposed method provides an efficient and scalable solution for intelligent industrial automation and collaborative robotic systems. The framework also improves adaptability in dynamic industrial environments where

activity patterns may vary between workers and tasks. By reducing the

requirement for manual data annotation, the proposed system lowers development cost and improves scalability for real-world deployment. The integration of unsupervised representation learning with probabilistic temporal modeling provides a reliable solution for continuous activity monitoring in smart manufacturing systems. The proposed approach can further support future industrial automation applications involving collaborative robotics, predictive maintenance and intelligent process optimization.

INDEX TERMS: Artificial Intelligence, Convolutional Neural Network, Deep Learning, Diabetic Retinopathy, Resolvable AI, Image Processing, Medical Image Analysis, Transfer Learning

1. INTRODUCTION

Electronic waste, commonly known as e-waste, has become one of the fastest growing waste streams across the world. The increasing use of electronic products and rapid technological advancements have resulted in a large quantity of discarded devices such as computers, hard disk drives, printers and mobile phones. In remanufacturing industries, these discarded products are disassembled to recover reusable components and valuable materials. However, most disassembly operations are still carried out manually because of the complexity involved in handling different product and tools.

Manual disassembly processes suffer from several limitations including low productivity, worker fatigue, inconsistent operation and high labor costs. To overcome these challenges, industries are gradually adopting human-robot collaboration systems. Collaborative robots can support workers during repetitive and physically demanding tasks. Nevertheless, successful collaboration requires robots to recognize.

Human Activity Recognition (HAR) has therefore become an important research area in industrial automation. Traditional activity recognition systems generally rely on supervised learning techniques that require large labeled datasets for training. Preparing labeled datasets for industrial activities is expensive.

To address these limitations, this research focuses on unsupervised learning methods for activity recognition in industrial disassembly environments. The proposed framework utilizes Variational Autoencoders to learn latent representations from unlabeled videos and Hidden Markov Models to identify temporal activity states. The objective is to develop an intelligent framework capable of recognizing human activities without heavy dependence on labeled data.

2. LITERATURE REVIEW:

Human activity recognition has gained significant attention in recent years in areas such as computer vision, machine learning and industrial automation. Earlier research mainly focused on traditional machine learning techniques where handcrafted features were extracted manually from video sequences. Algorithms such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Decision Trees were commonly used for recognizing human activities [4]. These methods mainly relied on motion information, body posture and object interaction features. Although these approaches produced reasonable results in controlled environments, their performance was not very effective in complex industrial situations where activities continuously change.

Later, with the growth of deep learning technologies, researchers started using Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for activity recognition tasks [5][6]. CNN models became useful for extracting spatial features automatically from video frames, while RNN and LSTM networks helped in understanding sequential and temporal relationships between activities [7]. Compared to traditional methods, deep learning approaches improved

recognition accuracy and reduced the need for manual feature engineering.

Even though deep learning methods achieved better performance, most of these approaches depended heavily on supervised learning techniques. In industrial environments, collecting and labeling large amounts of activity data is difficult, expensive and time consuming. Because of these limitations, recent research has started focusing more on unsupervised and self-supervised learning methods for activity recognition [8]. These approaches are capable of learning activity patterns directly from unlabeled data with less human effort.

Variational Autoencoders (VAE) have shown promising results in learning compressed latent representations from unlabeled video data [1]. Similarly, Hidden Markov Models (HMM) are widely used for sequential activity modeling and recognizing transitions between different activity states [2]. Some researchers have also combined deep learning models with probabilistic sequence modeling techniques to improve activity recognition performance in industrial environments [3].

Recent studies have further explored hybrid models that combine deep learning, clustering and temporal modeling approaches for continuous activity understanding [9]. Researchers have also investigated the use of sensor fusion, motion tracking and industrial video analysis techniques for improving activity recognition systems [10]. However, many existing systems still face problems related to scalability, computational complexity and adaptability to real industrial environments. Another major challenge observed in previous studies is the dependency on large annotated datasets, which are difficult to prepare in manufacturing industries. Therefore, unsupervised learning approaches are becoming more important because they provide a practical solution for learning useful activity patterns from unlabeled industrial video data.

Although considerable progress has been achieved in the field of activity recognition, challenges such as real-time implementation, reliable activity segmentation and adaptability to different industrial workstations still remain open research problems. The proposed framework attempts to address these issues using an efficient unsupervised learning architecture suitable for industrial disassembly environments.

3. PROPOSED SYSTEM:

3.1 System description:

The proposed system is an intelligent unsupervised human activity recognition framework designed for industrial disassembly environments. The framework mainly focuses on recognizing and analyzing human activities from video sequences during disassembly operations. The developed system is intended to support human-robot collaboration by enabling robots to understand worker activities accurately and efficiently without relying heavily on manually labeled datasets. The proposed framework uses advanced deep learning and probabilistic modeling techniques to recognize sequential human actions from industrial video data. The system combines a Variational Autoencoder (VAE) for unsupervised feature representation learning and a Hidden Markov Model (HMM) for temporal activity segmentation and classification. The Variational Autoencoder learns meaningful latent features

directly from unlabeled video frames, while the Hidden Markov Model identifies activity state transitions and recognizes sequential disassembly actions.

The framework processes activity videos through a structured workflow beginning from video acquisition and preprocessing to feature extraction, latent representation learning and activity recognition. The preprocessing stage standardizes video frames by performing resizing, normalization and noise reduction operations to improve recognition performance. The extracted latent features are then analyzed using probabilistic sequence modeling to recognize activities such as screw removal, tool handling, component separation and inspection tasks commonly performed during hard disk drive disassembly.

The proposed system is designed to reduce dependency on supervised learning methods that require extensive manual annotation. By learning directly from unlabeled video data, the framework provides a scalable and cost-effective solution for industrial environments where preparing labeled datasets is difficult and time consuming. The recognized activities can further assist collaborative robotic systems in understanding worker behavior and improving coordination during industrial operations.

The framework also aims to improve adaptability and efficiency in smart manufacturing environments. The proposed system can be extended for real-time industrial deployment where robots continuously observe worker activities and provide intelligent assistance during disassembly operations. The developed framework contributes toward the advancement of Industry 4.0 and intelligent automation technologies.

3.2 Functional Modules:

1. Video Input Module

This module is responsible for collecting industrial activity videos from cameras installed near the disassembly workstation. The module captures human activities performed during electronic product disassembly processes. It supports common video formats and ensures that the captured video data is suitable for further processing and analysis.

2. Video Preprocessing Module

The preprocessing module improves the quality and consistency of input video frames before feature extraction. Since industrial videos may contain variations in lighting conditions, background noise and camera angles, preprocessing is necessary to standardize the data. This module performs frame extraction, resizing, normalization and noise reduction operations. The preprocessing stage helps improve the efficiency and accuracy of activity recognition.

3. Feature Representation Learning Module

This module is the core component of the proposed framework where unsupervised feature learning is performed using a Variational Autoencoder (VAE). The VAE automatically learns meaningful latent representations from unlabeled video frames without requiring manual feature engineering. The encoder network compresses video frames into low-dimensional feature vectors, while the decoder reconstructs the original frames. The

learned latent representations capture important spatial and motion-related information associated with human activities.

4. Activity Segmentation Module

This module uses the Hidden Markov Model (HMM) to analyze sequential activity patterns and divide continuous disassembly operations into multiple activity states. The HMM models temporal dependencies between actions and identifies transitions occurring during disassembly tasks. This module helps recognize the order and duration of industrial activities more accurately.

5. Activity Recognition Module

The activity recognition module classifies disassembly activities based on the latent features and temporal patterns generated by previous modules. The framework recognizes industrial actions such as tool picking, screw removal, component separation and inspection processes. The recognized activity labels are generated automatically without depending heavily on manually annotated training data.

6. Performance Evaluation Module

This module evaluates the effectiveness of the proposed framework using performance metrics such as accuracy, precision, recall and F1-score. The recognized activities are compared with ground truth activity labels for validation purposes. The evaluation module helps analyze recognition performance and system reliability.

7. User Interface Module

The user interface module provides a simple and interactive environment for monitoring activity recognition results. Users can upload industrial activity videos, observe recognized activities and analyze system outputs. The interface is designed to support future real-time deployment in industrial human-robot collaborative systems.

3.3. System Workflow

The workflow of the proposed unsupervised human activity recognition framework begins with capturing industrial disassembly activity videos using cameras installed near the workstation. The recorded videos contain sequential human actions such as tool handling, screw removal, component separation and inspection activities performed during hard disk drive disassembly.

After video acquisition, the input videos are passed to the preprocessing stage. In this stage, video frames are extracted and standardized using operations such as resizing, normalization and noise reduction. The preprocessing stage improves the quality and consistency of input data, which helps increase recognition performance.

The preprocessed video frames are then provided to the Variational Autoencoder (VAE). The encoder network compresses the input frames into low-dimensional latent feature vectors that contain important spatial and motion-related

information. The decoder reconstructs the original frames from the latent representations, allowing the model to learn meaningful activity features without requiring labeled training data.

The extracted latent feature vectors are then passed to the Hidden Markov Model (HMM). The HMM analyzes temporal relationships between activities and identifies transitions between different activity states. This process helps divide continuous disassembly operations into meaningful sequential actions.

After temporal modeling, the framework performs activity recognition and classification. The system identifies industrial activities such as screw removal, tool picking, component handling and inspection operations based on learned activity patterns.

Finally, the recognized activity results are displayed through the user interface module. The system performance is evaluated using metrics such as accuracy, precision, recall and F1-score by comparing predicted activities with ground truth labels. The complete workflow provides an efficient and scalable solution for intelligent industrial human-robot collaboration systems.

The framework also includes a feedback and refinement mechanism that continuously improves activity recognition performance using previously learned activity patterns.

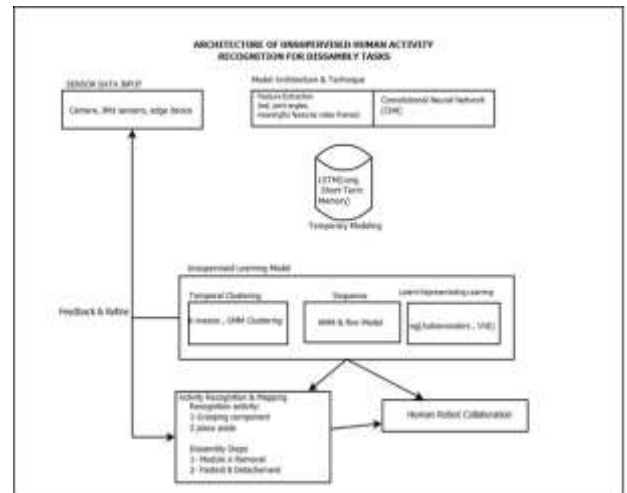


Fig 4.1 System Architecture

4. SYSTEM ARCHITECTURE

The architecture of the proposed unsupervised human activity recognition framework is designed to recognize and analyze industrial disassembly activities using sensor data, deep learning techniques and probabilistic sequence modeling. The framework mainly focuses on extracting meaningful activity patterns from industrial video sequences and identifying sequential human actions during disassembly operations.

The system begins with the Sensor Data Input layer where activity-related data is collected using cameras, IMU sensors and edge devices installed near the industrial workstation. The collected video and sensor data are then provided to the feature extraction stage. In this stage, important spatial and motion-related information such as joint angles, movement patterns and video frame features are extracted from the input data.

After feature extraction, the data is processed using Convolutional Neural Networks (CNNs) for learning important visual representations from activity frames. The extracted features are then forwarded to the temporal modeling stage where Long Short-Term Memory (LSTM) networks are used for analyzing sequential activity dependencies over time. The temporal modeling stage helps the framework understand continuous industrial activities and worker motion behavior.

The architecture also includes an unsupervised learning module consisting of temporal clustering, sequence modeling and latent representation learning components. Clustering methods such as K-Means and Gaussian Mixture Models (GMM) are used to group similar activity patterns. Hidden Markov Models (HMM) and Recurrent Neural Networks (RNN) are used for sequence modeling and activity state transition analysis. Autoencoders and Variational Autoencoders (VAE) are applied for latent representation learning from unlabeled activity data.

Finally, the recognized activities are mapped into industrial disassembly operations such as component grasping, placement and module removal tasks. The recognized activities assist human-robot collaboration systems in understanding worker actions and improving coordination during industrial operations.

1. Sensor Data Input Layer

This layer is responsible for collecting activity-related data from the industrial environment using cameras, IMU sensors and edge devices installed near the workstation. The cameras capture worker movements and disassembly operations in video form, while sensors record motion and positional information. The collected data acts as the primary input for the proposed framework and helps monitor industrial activities continuously during disassembly tasks.

2. Feature Extraction Layer

The feature extraction layer processes the collected raw video and sensor data to identify important activity-related information. Features such as hand movement, body posture, tool usage, motion patterns and object interaction are extracted from video frames. This layer helps remove unnecessary information and provides meaningful activity features that improve the learning and recognition capability of the framework.

3. Convolutional Neural Network (CNN) Layer

The CNN layer is used for automatic visual feature learning from industrial video frames. It analyzes spatial information and identifies important visual patterns related to worker activities and disassembly operations. The CNN helps detect object positions, movement structures and interaction patterns between workers and tools. This layer reduces the need for manual feature engineering and improves recognition performance.

4. Temporal Modeling Layer (LSTM)

The Long Short-Term Memory (LSTM) layer is responsible for analyzing sequential activity behavior over time. Since industrial

disassembly tasks involve continuous and ordered operations, the LSTM helps the framework understand the relationship between consecutive actions. It recognizes activity sequences such as tool picking, screw removal and component separation by learning temporal dependencies between activities.

5. Unsupervised Learning Layer

This layer performs activity learning without depending heavily on manually labeled datasets. It uses unsupervised learning techniques such as clustering and sequence modeling to identify hidden activity patterns directly from industrial video sequences. The layer improves scalability of the framework and reduces the time and effort required for manual activity annotation in industrial environments.

6. Latent Representation Learning Layer

The latent representation learning layer uses Autoencoders and Variational Autoencoders (VAE) to generate compressed feature representations from unlabeled industrial video data. These latent features contain important information related to worker motion, hand movement and activity behavior. This layer helps the framework learn complex activity patterns efficiently while reducing the dimensionality of input data.

7. Activity Recognition and Mapping Layer

This layer identifies industrial disassembly activities based on learned activity patterns and temporal relationships. Activities such as grasping components, placing objects, removing screws and module separation are recognized and mapped into meaningful industrial operations. The recognized activity information can be further used for monitoring worker performance and improving industrial process efficiency.

8. Human-Robot Collaboration and Feedback Layer

This layer allows collaborative robotic systems to understand worker activities and provide intelligent support during disassembly operations. The recognized activity information helps robots coordinate safely and efficiently with workers in real time. The feedback mechanism also continuously improves recognition performance by refining activity predictions based on previously learned patterns and industrial task behavior.

4.1 DFD Diagram:

4.1.1 Level 0 DFD:

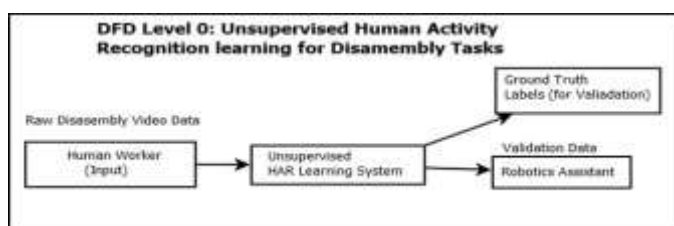


Fig 4.2 Level 0 Data Flow Diagram

The Level 0 DFD represents the overall workflow of the proposed Unsupervised Human Activity Recognition system for disassembly tasks. The process starts with the human worker performing industrial disassembly operations, where raw disassembly video data is captured as the input to the system. The collected video data is provided to the Unsupervised HAR Learning System, which automatically learns and recognizes human activity patterns from the industrial video sequences. The recognized activities are validated using Ground Truth Labels to evaluate system accuracy and performance. The processed validation data is then supplied to the Robotics Assistant system to support human-robot collaboration during industrial disassembly operations. This diagram mainly shows the interaction between the worker, activity recognition system, validation process and robotic assistance system in a simplified form.

4.1.2 Level 1 DFD:

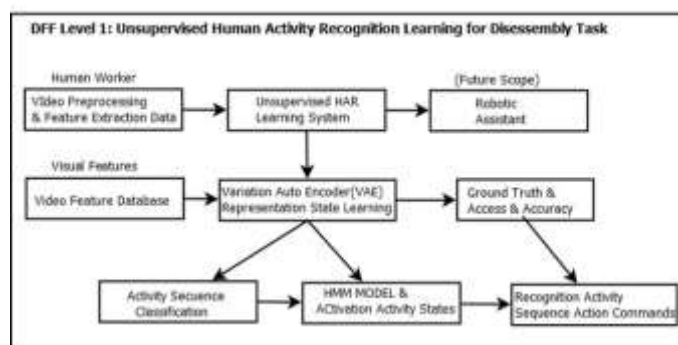


Fig 4.3 Level 1 Data Flow Diagram

The Level 1 DFD represents the detailed workflow of the proposed Unsupervised Human Activity Recognition framework for disassembly tasks. The process begins with video preprocessing and feature extraction from industrial activity videos captured during disassembly operations performed by the human worker. The extracted visual features are stored in the video feature database and provided to the Variational Autoencoder (VAE) for representation learning. The VAE learns compressed latent representations from the activity data and forwards the learned features for activity sequence classification and temporal activity state analysis. The Hidden Markov Model (HMM) is then used to identify sequential activity states and recognize continuous industrial actions. The recognized activity sequences are validated using ground truth data to measure recognition accuracy and system performance. Finally, the generated recognition activity sequence commands are provided to the robotic assistant system to support intelligent human-robot collaboration during industrial disassembly operations.

5. METHODOLOGY

The proposed framework follows a systematic methodology for recognizing human activities during industrial disassembly tasks using unsupervised learning techniques. The methodology is designed to reduce dependency on manually labeled datasets while improving activity recognition performance in industrial environments.

The first stage of the methodology involves data collection and video acquisition. Industrial activity videos are collected from disassembly workstations where workers perform hard disk drive dismantling operations. The recorded videos contain sequential human activities such as tool handling, screw removal, component separation and inspection processes. These activity videos are used as the input dataset for the proposed framework.

The second stage is video preprocessing and frame extraction. Since industrial video data may contain variations in illumination, background conditions and camera angles, preprocessing is necessary to improve data quality and consistency. In this stage, videos are divided into frames and preprocessing operations such as resizing, normalization and noise reduction are applied. The preprocessing stage helps improve feature learning and recognition accuracy.

The third stage involves unsupervised feature representation learning using the Variational Autoencoder (VAE). The encoder network compresses high-dimensional video frames into low-dimensional latent feature vectors that capture important spatial and motion-related activity information. The decoder reconstructs the original frames from the latent representations. Through this reconstruction process, the VAE learns meaningful activity features automatically without relying on labeled training data.

The fourth stage focuses on temporal activity modeling using the Hidden Markov Model (HMM). The extracted latent feature vectors generated by the VAE are provided as input to the HMM. The Hidden Markov Model analyzes sequential activity patterns and identifies transitions between different activity states. This stage helps divide continuous disassembly operations into meaningful action sequences and improves recognition of temporal activity dependencies.

The final stage is activity classification and performance evaluation. The framework recognizes industrial activities such as screw removal, component handling and inspection operations based on learned temporal patterns. The performance of the proposed system is evaluated using metrics such as accuracy, precision, recall and F1-score. The predicted activities are compared with ground truth labels to measure the effectiveness and reliability of the proposed framework in industrial human-robot collaboration environments.

6. IMPLEMENTATION

The proposed unsupervised human activity recognition framework is implemented using Python programming language along with deep learning and computer vision libraries. The complete system is developed in a modular manner so that each stage of the framework can be processed efficiently. Libraries such as TensorFlow, Keras and OpenCV are used for video processing, feature learning and activity recognition tasks. The implementation mainly focuses on recognizing industrial disassembly activities from video sequences without depending heavily on manually labeled data. The implementation process begins with industrial video data collection. Activity videos are recorded from disassembly workstations where workers perform hard disk drive dismantling operations. These videos contain multiple sequential activities such as tool handling, screw removal,

component separation and inspection processes. The collected videos are stored and organized properly for preprocessing and training purposes. Different video samples are captured under varying environmental conditions to improve the robustness of the framework.

In the preprocessing stage, the recorded videos are divided into image frames using OpenCV libraries. The extracted frames are resized into fixed dimensions and normalized to maintain consistency throughout the dataset. Noise reduction and frame enhancement operations are also performed to improve video quality. Proper preprocessing helps improve feature extraction performance and reduces unnecessary variations caused by lighting conditions or camera movement.

After preprocessing, the Variational Autoencoder (VAE) model is implemented for unsupervised feature representation learning. The encoder network converts input video frames into compressed latent feature vectors, while the decoder reconstructs the original frames from these latent representations. The VAE model is trained using unlabeled industrial video data so that it can automatically learn important activity-related features without manual annotation. During training, reconstruction loss and latent distribution regularization are used to optimize model performance.

The latent feature vectors generated by the VAE are then provided to the Hidden Markov Model (HMM) for temporal activity modeling. The HMM analyzes sequential activity patterns and identifies transitions between different disassembly states. Since industrial tasks generally follow continuous step-by-step operations, the HMM effectively captures temporal dependencies between activities and improves activity segmentation accuracy.

After temporal modeling, the framework performs activity recognition and classification based on learned activity patterns. The system identifies industrial actions such as screw removal, component handling and tool usage operations. The recognized activities are compared with actual activity labels to evaluate framework performance. Metrics such as accuracy, precision, recall and F1-score are used for evaluation purposes.

The developed framework is tested using hard disk drive disassembly activity videos collected from industrial-like environments. Experimental results demonstrate that the framework performs efficiently in recognizing sequential industrial activities while reducing dependency on labeled training data. The implementation also shows that combining unsupervised feature learning with probabilistic sequence modeling provides better scalability and adaptability for future human-robot collaboration systems in smart manufacturing environments.

7. RESULTS AND DISCUSSION

The proposed unsupervised human activity recognition framework was evaluated using industrial disassembly activity videos involving hard disk drive dismantling operations. The framework was tested on multiple sequential activities such as tool handling, screw removal, component separation and inspection processes. Experimental analysis was carried out to evaluate the effectiveness of the proposed system in recognizing industrial activities from unlabeled video data.

The Variational Autoencoder (VAE) successfully learned meaningful latent feature representations from industrial video sequences without requiring manually labeled datasets. The generated latent vectors contained important spatial and motion-related information associated with worker activities. The Hidden Markov Model (HMM) effectively analyzed temporal activity transitions and segmented continuous disassembly operations into meaningful activity states. The combination of VAE and HMM improved the ability of the framework to recognize sequential industrial activities accurately.

The proposed framework achieved an average activity recognition accuracy of approximately 91.52%, which was higher compared to several conventional machine learning approaches used for industrial activity recognition. The preprocessing stage helped improve consistency in video data, which positively affected feature extraction and overall recognition performance. The framework also demonstrated stable recognition performance during continuous industrial operations.

Compared to supervised learning approaches, the proposed system significantly reduced dependency on manually labeled training data. This is an important advantage because preparing labeled industrial activity datasets requires extensive human effort and time. The developed framework provides a more scalable and cost-effective solution for industrial environments where large amounts of unlabeled video data are available.

The experimental results also indicate that probabilistic temporal modeling using Hidden Markov Models improves activity segmentation and sequential action understanding. Industrial disassembly tasks generally involve ordered workflows, and the HMM effectively captured transitions between consecutive activities. This improved recognition stability during continuous operations and reduced misclassification between similar activities.

The obtained results demonstrate that the proposed framework has strong potential for future implementation in industrial human-robot collaboration systems and smart manufacturing environments. Although the framework achieved reliable performance, recognition accuracy may decrease under poor lighting conditions or rapid worker movement. Future improvements involving advanced deep learning architectures and multimodal sensor integration may further improve recognition performance and real-time deployment capability.

8. FUTURE WORK

The proposed framework can be further improved by integrating advanced deep learning architectures such as transformer-based models and self-supervised learning techniques for better feature extraction and activity understanding. Future work may also include real-time deployment in industrial robotic systems for continuous human-robot collaboration. In addition, integrating multiple sensor inputs such as depth cameras and motion sensors can further improve recognition accuracy and system reliability in complex industrial environments.

The framework can also be extended for recognizing activities in different industrial applications such as assembly lines, maintenance operations and warehouse automation systems. Further improvements in activity segmentation and adaptive learning methods may help the system perform efficiently under varying workstation conditions and worker behavior patterns.

These advancements can contribute toward the development of more intelligent and fully automated smart manufacturing environments.

9. CONCLUSION

In this research work, an unsupervised human activity recognition framework for industrial disassembly tasks has been proposed using Variational Autoencoders (VAE) and Hidden Markov Models (HMM). The framework was developed to recognize sequential human activities from industrial video data without relying heavily on manually labeled datasets. The proposed system focuses on improving activity understanding in human-robot collaboration environments commonly found in remanufacturing industries.

The Variational Autoencoder was used for learning meaningful latent feature representations from unlabeled video sequences, while the Hidden Markov Model was used for temporal activity segmentation and sequential activity recognition. The combination of deep representation learning and probabilistic temporal modeling improved the overall recognition performance of the framework. Experimental analysis demonstrated that the proposed system achieved reliable activity recognition accuracy during hard disk drive disassembly operations.

The developed framework also reduced the dependency on supervised learning approaches, which generally require extensive manual data annotation and large labeled datasets. This makes the proposed system more scalable and practical for real industrial environments where collecting labeled activity data is difficult and time consuming.

Overall, the proposed framework provides an effective solution for intelligent industrial activity recognition and has strong potential for future implementation in smart manufacturing and Industry 4.0 applications. The developed system can contribute toward improving productivity, automation and coordination in future human-robot collaborative industrial environments.

10. REFERENCES

1. Kingma, D. P., and Welling, M., "Auto-Encoding Variational Bayes," International Conference on Learning Representations (ICLR), 2014.
2. Rabiner, L. R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.
3. Wang, J., Chen, Y., Hao, S., Peng, X., and Hu, L., "Deep Learning for Sensor-based Activity Recognition: A Survey," Pattern Recognition Letters, vol. 119, pp. 3–11, 2019.
4. Aggarwal, J. K., and Ryoo, M. S., "Human Activity Analysis: A Review," ACM Computing Surveys, vol. 43, no. 3, pp. 1–43, 2011.
5. Simonyan, K., and Zisserman, A., "Two-Stream Convolutional Networks for Action Recognition in Videos," Advances in Neural Information Processing Systems (NIPS), 2014.

6. Carreira, J., and Zisserman, A., “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
7. Donahue, J., Hendricks, L. A., Guadarrama, S., et al., “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
8. Srivastava, N., Mansimov, E., and Salakhutdinov, R., “Unsupervised Learning of Video Representations using LSTMs,” International Conference on Machine Learning (ICML), 2015.
9. Ji, S., Xu, W., Yang, M., and Yu, K., “3D Convolutional Neural Networks for Human Action Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221–231, 2013.
10. Koppula, H. S., Gupta, R., and Saxena, A., “Learning Human Activities and Object Affordances from RGB-D Videos,” International Journal of Robotics Research (IJRR), vol. 32, no. 8, pp. 951–970, 2013.

Authors:

Mr. RUSHIKESH MURLIDHAR CHORMALE is currently pursuing a Bachelor's Degree in Computer Science & Engineering at Ashokrao Mane Group Of Institution, vathar. He is publishing this paper under the esteemed guidance of Mr. Anil B. Desai.



Mr. NIKHIL ARJUN GHARPANKAR is currently pursuing a Bachelor's Degree in Computer Science & Engineering At Ashokrao Mane Group of Institution, Vathar. He is publishing this paper under the esteemed guidance of Mr. Anil B. Desai.



Mr. ONKAR SURESH KHARUDE is currently pursuing a Bachelor's Degree in Computer Science & Engineering at Ashokrao Mane Group Of Institution, vathar.. He is publishing this paper under the esteemed guidance of Mr Anil B. Desai.



Mr. SAHIL SANJAY KAGINKAR is currently pursuing a Bachelor's Degree in Computer Science & Engineering at Ashokrao Mane Group Of Institution, vathar. He is publishing this paper under the esteemed guidance of Mr. Anil B. Desai.



Mr. PRAVIN RAJENDRA KAMBLE is currently pursuing a Bachelor's Degree in Computer Science & Engineering At Ashokrao Mane Group of Institution, Vathar. He is publishing this paper under the esteemed guidance of Mr. Anil B. Desai.



Mr. Anil B. Desai is currently working as an Assistant Professor in Computer science and engineering At Ashokrao Mane Group of Institution, Vathar. He has been guiding and mentoring students in various academic and research activities. He has provided his valuable guidance for the successful completion and publication of this paper.

