Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/16/20230883

Behavior recognition technology research and application analysis

Kaixuan Wei

Hefei 168 Middle School Department, Hefei, 230022, China

kyle985wkx@mail.sdufe.edu.cn

Abstract. This paper provides a comprehensive overview of the current state of behavior recognition technology research and its applications in computer vision. Firstly, it discusses the fundamental concepts and categorization methods employed in behavior recognition technology, such as multimodality, double-sided depth photos, bone key points, and RGB data. These techniques enable the recognition and analysis of various human behaviors with high accuracy and precision. Furthermore, this paper highlights the vast potential of behavior recognition technology in several fields, including safety and education. In safety settings, behavior recognition technology can assist managers in identifying abnormal behaviors and enhancing safety precautions. In educational settings, behavior recognition technology can help teachers gain insight into student learning levels and enhance their teaching efficiency. Additionally, this technology can be used to identify patterns of behaviors that might indicate a student is struggling or needs extra support. The paper concludes with a summary of the current state of behavior recognition research and suggests areas for further investigation. One potential area for future research is the development of more accurate and efficient recognition models. Additionally, exploring the ethical implications and privacy concerns of behavior recognition technology is also essential. Overall, this paper emphasizes the immense potential of behavior recognition technology in various fields and encourages further research to realize its full potential. By leveraging the power of computer vision, we can gain valuable insights into human behavior that could have far-reaching implications for safety, education, and other areas of our lives

Keywords. action recognition, computer vision, double-sided depth images, skeletal keypoints, multimodality.

1. Introduction

Computer vision technology has rapidly evolved alongside artificial intelligence, becoming an essential tool in various industries [1]. The development of behavior recognition technology has become a significant research area in computer vision, allowing for the recognition and assessment of human behavior through the processing and analysis of video, images, and other data [2]. Behavior recognition technology is widely used in education, security, and healthcare industries [3].

In the classroom, behavior recognition technology can aid teachers in understanding students' learning state, enhancing instruction quality and effectiveness. Additionally, it can assist security personnel in spotting unusual behavior, improving security measures [4]. In the medical industry,

DOI: 10.54254/2755-2721/16/20230883

behavior recognition technology can improve diagnosis and treatment precision by assisting clinicians in recognizing patients' illnesses and symptoms [5].

However, behavior recognition technology faces numerous challenges, such as high complexity and real-time demands, requiring fast and accurate identification of human behaviors [6]. Typical behavior recognition techniques may struggle to produce high-precision findings due to challenging feature extraction and reliance on high-quality data [7]. Although deep learning methods have gained popularity in action recognition, they still face limitations in recognizing patterns and functioning effectively in real-world scenarios, impacted by factors such as training set quantity, data quality, and algorithm complexity [8].

Therefore, current behavior recognition technology research aims to address the challenge of producing effective, high-precision, and real-time behavior identification technology [9]. Academics work tirelessly to advance behavior detection technology via improved data accuracy, algorithm refinement, and real-time performance enhancement. This article summarizes the recent state-of-the-art behavior recognition technology and application based on deep learning, discusses its limitations, and examines the prospects for its real-world applications, aiming to provide a reference for future research and practical implementation [10].

2. Identify related technologies

2.1. Based on rgb data

Behavior recognition is a prominent research area in computer vision, with behavior recognition based on RGB data being a popular approach. This method consists of two categories: the manual feature approach and the machine learning approach.

The manual feature approach includes action representation methods based on spatiotemporal volume, spatiotemporal interest points, and skeletal joint trajectory. These methods primarily involve extracting motion and spatiotemporal changes in human activities [9]. For example, spatiotemporal volume uses 3D spatiotemporal templates for action recognition, requiring the construction and encoding of motion data. Spatiotemporal interest points express actions using key locations with clear motion changes, such as 3D Harris spatiotemporal feature points. Action representation based on the trajectory of skeletal joints tracks bone movements or extracts key points from an RGB image to classify actions based on trajectory and key points.

The machine learning approach achieves behavior recognition through feature extraction and classification of RGB data. Both deep learning networks and conventional manual feature extraction techniques can be used for feature extraction, and common classification techniques include SVM and decision trees. While the RGB-based behavior recognition method has the benefits of being inexpensive and easy to acquire, it lacks robustness to appearance changes and faces difficulty handling situations where recognition targets and backgrounds have similar colors and textures.

To effectively achieve behavior recognition in practical applications, it is necessary to carefully evaluate the benefits and drawbacks of different approaches.

2.2. Based on double-sided depth images

Double-sided depth images are being utilized more frequently in behavior identification technology based on depth images [5]. Depth photos are more accurate in capturing the depth information of objects in a scene than typical RGB images because they are unaffected by environmental elements like lighting. This increases the precision of behavior recognition.

Currently, there is a lot of study and use of behavior recognition technologies based on double-sided depth pictures. Li et al., for instance, developed the idea of the Action Graph, which realizes behavior modeling through the use of depth pictures for human body contour edge extraction and bag-of-words model encoding. Yang et al. combined the HOG features for recognition after projecting the depth image into three orthogonal planes. Jalal et al. integrated PCA dimension reduction and hidden Markov model

DOI: 10.54254/2755-2721/16/20230883

for behavior identification, using R transformation to convert the contour of the depth image into a more compact feature expression.

Additionally, double-sided depth photographs have been extensively used in real-world applications for the behavior identification technology. For instance, behavior identification technology based on double-sided depth photographs can enable real-time monitoring and analysis of staff activity trajectories and enhance security prevention capabilities in the field of intelligent monitoring. Robots can interpret and evaluate human behavior with the aid of behavior recognition technology based on double-sided depth pictures, which enhances the robot's interactive skills and service level.

As a result, the technology for behavior recognition based on double-sided depth photographs is not only significant for theoretical study but also has many potential applications in real-world settings.

2.3. Based on skeletal keypoints

Bone data, a brand-new class of multimodal data, has particular advantages in the identification of human behavior. Because bone data more properly expresses the dynamic properties of the human skeleton, it can better reflect human behavior when compared to typical 2D RGB images and depth data. Additionally, bone data has the advantages of being relatively small and having a quick calculation time. As a result, bone data is a prospective research area for action recognition.

Since the movement of human bones can distinguish between a wide variety of movements, bone data contains rich spatiotemporal information, and joint nodes and their adjacent nodes have a strong correlation, skeleton data can not only find rich human structure in the same frame. A substantial link exists between frames and information. These qualities offer a solid basis for the use of bone data.

The key to using bone data effectively is learning how to extract its properties and apply them to the task of identifying behavior. A typical technique at the moment is to project all of the skeleton points in the entire human behavior, divide the image into sub-grids, extract the frequency of the projected points, and then normalize them. In order to accurately characterize human behavior, this method can be utilized to define the cumulative projection distribution of human skeleton joint points and express the relationship of all pertinent nodes in space. The approach still has to be enhanced for the identification of identical activities with marginal interclass variations [3].

Technology for recognizing behavior based on skeleton key points performs well in various real-world application contexts. For instance, this technology can be applied to video surveillance to recognize and track human behavior, increasing the effectiveness and precision of the systems. The user experience can be improved by using virtual reality technology in interactive programs and games. With regard to athletic competition, this technology can be utilized to analyze and assess the movement traits of players and enhance the effectiveness of training. As a result, there are numerous potential applications for the behavior identification system based on skeletal key points.

2.4. Based on multimodal

The difficult issue of multimodal behavior detection necessitates the merging of input from several sensors to produce more precise and comprehensive information. Visual and non-visual modalities make up the majority of multimodal data. The visual modality typically consists of the RGB image, depth image, skeleton point sequence, etc. among them. These data come from several sources, so each one has its own own information and features, but there are also overlaps and complementarities. Therefore, the fundamental problem of multimodal fusion is how to extract and fuse pertinent complimentary information from multimodal data.

Two popular techniques for multimodal fusion are early fusion and late fusion. Early fusion involves combining data from several modalities at the network's input layer. Feature map addition, feature map stacking, convolution, multi-modal fusion structure search, and attention mechanism are some of the frequently utilized techniques [4]. This technique can combine several levels of semantic data to get superior outcomes. However, early fusion also faces a number of challenges, such as figuring out how to create multi-modal feature extraction networks and how to assess each modality's feature relevance.

DOI: 10.54254/2755-2721/16/20230883

The goal of late fusion is to combine the output of various modalities at the network's output layer. Predicting score averaging, maximum value, multiplication, addition, attention mechanism, and knowledge distillation are common techniques. Although the late fusion method is straightforward and simple to use, it only achieves modest improvement since it overlooks the association among multimodal data. Therefore, in real-world applications, late fusion typically performs less well than early fusion.

Effectively fusing data from several modalities is the key to multimodal fusion. Adopting an attention mechanism to combine many aspects early on while learning the relative importance of each modality's data is a successful strategy. To optimize the effectiveness of the fused features, another strategy is to employ multimodal fusion structure search to automatically look for fusion structures. The effect of multimodal fusion is enhanced in late fusion by knowledge distillation, which can transmit knowledge from one model to another during learning.

3. Behavior recognition application

3.1. Classroom behavior recognition

A cutting-edge application of artificial intelligence is classroom behavior identification technology. Its goal is to continuously track and assess how pupils are learning in order to enhance the effectiveness of instruction. Through image processing, deep learning, and other technical methods, this technology can automatically categorize and recognize students' behavior in the classroom, allowing teachers to monitor the learning environment and classroom dynamics and provide more accurate data support for the classroom [6].

Some high-tech companies have started creating and marketing classroom behavior recognition system products in recent years, like the "Smart Classroom Behavior Management System," and have started testing them in some schools. The system can use smart cameras, sensors, and other technology to gather, analyze, and classify student behavior data in real time while they are in the classroom. Teachers can better carry out instructional activities by learning information such as student engagement, attentiveness, emotional condition, etc. through systematic data statistics and analysis.

However, there are still certain difficulties and problems with this technology. The first is the problem in compiling statistics on student conduct in the classroom. Because there is such a wide variety in student behavior in the classroom, it is difficult to create and identify datasets. Second, in order to guarantee the accuracy and applicability of the data, this technology must strike a balance between the two in order to handle the issue of real-time and accurate multi-target detection. The accuracy of target behavior recognition in real-world scenarios must also be improved, for example.

Future algorithm process and model structure improvements will be required to address these issues and increase the effectiveness and precision of detection. Constructing student standard behavior data sets, establishing behavior definition standards, and improving the precision and toleration of behavior characteristic definitions are all necessary at the same time. To further protect students' personal privacy and information security, policies including those pertaining to privacy protection and data security must be improved.

3.2. Safe behavior recognition

Behavior recognition technology can be utilized in the area of safety management to detect and identify whether employees follow safety standards and operational procedures, thereby significantly enhancing workplace safety and productivity.

In the topic of safety management, behavior recognition technology is now a research hotspot, and related research spans a variety of sectors and fields. Helmet use is a crucial safety requirement in hydropower plant safety management. As a result, it has become crucial to conduct research on how to use behavior recognition technology to detect and determine if workers are wearing helmets properly.

The majority of traditional behavior recognition technology focuses on labor-intensive feature extraction and classification, yet this method has weaknesses like low recognition accuracy and efficiency. For the purpose of identifying behavior, the single frame multi-target detection algorithm

DOI: 10.54254/2755-2721/16/20230883

(SSD) is a popular deep learning model. Deep learning models are primarily used for behavior detection in the current behavior recognition technology due to advancements in computer vision and deep learning. The SSD model can recognize and identify people wearing hard hats by preprocessing pictures and extracting attributes, improving the effectiveness and precision of safety management.

With the advancement of neural network architecture in recent years, SENet network has become a popular network structure in the areas of image recognition and behavior recognition. The accuracy and stability of the SSD model, which is based on the SENet network structure, may be improved more successfully, and the accuracy and effectiveness of recognizing staff safety behavior can also be increased.

4. Epilogue

With the aid of computer vision and machine learning algorithms, behavior recognition technology analyzes and recognizes the traits of human behavior to automatically monitor and evaluate behaviors. One of the hottest areas in the disciplines of artificial intelligence and computer vision in recent years has been behavior recognition technology, which is finding increasing use in a variety of fields.

The study of behavior recognition technology at this time primarily concentrates on aspects based on RGB data, based on double-sided depth photographs, based on bone key points, and based on multimodality. Among them, behavior recognition based on RGB data analyzes and recognizes behavior using color photos taken with regular cameras. Although light and background interference can be avoided by using the depth camera's double-sided image for behavior recognition, doing so comes at a high equipment cost. Behavior recognition based on bone key points uses the bone key points that the depth camera has collected in order to perform behavior recognition. The benefits of recognition include high accuracy and low cost; multi-modal behavior recognition combines input from several sensors to increase the reliability and accuracy of behavior recognition. The disciplines of security and education have both made extensive use of behavior recognition technologies in addition to the technological studies mentioned above. In order to provide better educational services, behavior recognition technology can assist teachers in understanding students' learning levels and behavioral patterns. In the area of security, behavior recognition technology can assist security personnel in monitoring and averting security incidents as well as enhancing overall public security.

Behavior recognition technology is anticipated to be used in more industries in the future, including autonomous driving, smart homes, and health management, thanks to the ongoing development of technologies like computer vision and artificial intelligence. Behavior recognition technology can be used in smart homes to automate the operation of household appliances based on family members' propensities for certain behaviors, and it can also be used in automatic driving systems to assist automobiles in recognizing pedestrians, other vehicles, and surrounding vehicles. Furthermore, the medical industry makes extensive use of behavior recognition technologies. A more thorough and precise health evaluation can be given to medical professionals by combining computer vision technologies and sensors, for instance, to track and evaluate patient activity. Additionally, the care and supervision of populations including the elderly, the disabled, and kids can be done using behavior recognition technology. Their quality of life can be improved by continuously monitoring their behavior patterns and identifying aberrant circumstances.

In general, behavior recognition technology is utilized extensively in a variety of industries and has greatly improved people's lives in terms of ease and security. The range of application scenarios for behavior recognition technology will expand in the future as a result of the ongoing advancement and development of technology, improving people's lives in terms of ease and security.

5. Conclusion

In conclusion, behavior recognition technology has shown remarkable progress and potential in various fields, including safety and education. This paper has discussed the fundamental concepts and categorization methods used in behavior recognition technology, such as multimodality, double-sided depth photos, bone key points, and RGB data. These techniques enable accurate and precise recognition

DOI: 10.54254/2755-2721/16/20230883

and analysis of various human behaviors. Furthermore, behavior recognition technology can assist managers in identifying abnormal behaviors and enhancing safety precautions in safety settings. In educational settings, this technology can help teachers gain insight into student learning levels and enhance their teaching efficiency. It can also be used to identify patterns of behaviors that might indicate a student needs extra support or is struggling. The future of behavior recognition technology holds great potential for more accurate and efficient recognition models, as well as exploring the ethical implications and privacy concerns of this technology. As we continue to leverage the power of computer vision, we can gain valuable insights into human behavior that could have far-reaching implications for various areas of our lives. Therefore, further research is necessary to fully realize the potential of behavior recognition technology. By combining the advancements in this technology with various fields, we can enhance safety measures, improve education, and gain valuable insights into human behavior. This paper encourages researchers to explore these opportunities and challenges and contribute to the continued development of behavior recognition technology.

References

[1] Wang P. Research on sports training action recognition based on deep learning[J]. Scientific Programming, 2021, 2021: 1-8.

[2] Pareek P, Thakkar A. A survey on video-based human action recognition: recent updates, datasets, challenges, and applications[J]. Artificial Intelligence Review, 2021, 54: 2259-2322.

[3] Sarkar A, Banerjee A, Singh P K, et al. 3D Human Action Recognition: Through the eyes of researchers[J]. Expert Systems with Applications, 2022: 116424.

[4] Cheng M, Yuan H, Wang Q, et al. Application of deep learning in sheep behaviors recognition and influence analysis of training data characteristics on the recognition effect[J]. Computers and Electronics in Agriculture, 2022, 198: 107010.

[5] Tsai J K, Hsu C C, Wang W Y, et al. Deep learning-based real-time multiple-person action recognition system[J]. Sensors, 2020, 20(17): 4758.

[6] Segalin C, Williams J, Karigo T, et al. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice[J]. Elife, 2021, 10: e63720.

[7] Al-Faris M, Chiverton J, Ndzi D, et al. A review on computer vision-based methods for human action recognition[J]. Journal of imaging, 2020, 6(6): 46.

[8] Yang Q, Xiao D. A review of video-based pig behavior recognition[J]. Applied Animal Behaviour Science, 2020, 233: 105146.

[9] Shaikh M B, Chai D. Rgb-d data-based action recognition: A review[J]. Sensors, 2021, 21(12): 4246.

[10] Degardin B, Proença H. Human behavior analysis: a survey on action recognition[J]. Applied Sciences, 2021, 11(18): 8324.