Design and Implementation of Digital Media Technology Experimental System Based on Machine Learning

Yancong Su

School of Design Art, Xiamen University of Technology, Xiamen, China

Keywords: Machine Learning; Digital Media Technology; Human-Computer Interaction; Robots

Abstract: Digital media technology is an open platform, which mainly includes scene design, role image design, game programming, multimedia post-processing and human-computer interaction technology. In this paper, machine learning is used to study the recognition and Simulation of human motion by humanoid robots, so as to enhance human-computer interaction experience. Four tasks have been done, including moving object detection and tracking, motion feature representation, motion recognition and robot imitation, with emphasis on the use of machine learning methods, extraction of human body features, and research and implementation of motion recognition and simulation.

1. Introduction

In human communication, vision, hearing, olfaction, touch and so on are all important ways of expression, especially visual information. Through the images received by the eyes, we can determine the position, image, facial expression, limb movement and other characteristics of others, and provide context deduction information for ourselves, that is, language provides direct information, while visual help. Helping to identify scenarios, correspondingly, computer vision is indispensable in human-computer interaction [1]. Computer vision is based on the image processing of human vision, which enables the computer to have “vision”, that is, the ability to obtain corresponding information through the processing of images and videos. As mentioned earlier, when human beings communicate, they express their emotions through language and action. Speech research has developed more comprehensively and tends to be mature in commercial applications, but there are some gaps in action research [2]. Human body movement is a language to help human express their emotions and thoughts and strengthen human communication. In order to make humanoid robots more widely used and “smarter”, humanoid robots imitate human body movements as a way of human-computer interaction, making humanoid robots have the natural and intuitive interaction of human habits. Force obviously means a lot. In addition, human motion recognition is also an important research direction in the field of computer vision, and has broad application prospects [3].

2. Machine Learning Theory

Machine learning is a science that studies how to make machine simulate human learning activities to continuously improve its performance. It is an important subject in the field of artificial intelligence, a fundamental way to make machine intelligent, and also a research hotspot in the field of IT and Internet. In essence, machine learning is the integration of many disciplines, involving biology, psychology, probability and statistics, computer science, convex analysis, algorithm complexity and so on [4]. A machine learning working mechanism basically includes four parts: environment, learning components, knowledge base and execution components. The environment provides external information for the system. After the learning component gets the information, it analyses and modifies the knowledge base [5]. The knowledge base represents the performance standard of the system, executes the components to process the tasks according to the performance standard, and evaluates the completion of the tasks. The evaluation results are fed back to the learning component [6]. The learning component judges whether or not according to the results.

  DOI: 10.25236/scmc.2019.035

The knowledge base needs to be modified.

Since the 1970s, machine learning has been applied in image recognition, artificial intelligence, expert system, license plate recognition, data analysis [7]. With the development of machine learning theory, more and more scholars have devoted themselves to the research of machine learning, and their applications are more and more extensive, such as data mining, speech recognition, natural language processing, video automatic recognition, target tracking detection, face recognition, credit card fraud detection, etc. [8]. Until now, in the era of big data, many commercial companies have foreseen the broad application prospects of machine learning, invested a lot of manpower and material resources in theoretical research and product development, cloud computing, machine translation, network security, search engine, stock exchange, medical analysis, smart home, shape identification, customer segmentation, electricity. Commercial product recommendation, etc. At present, Microsoft, Amazon and Google have released Microsoft Azure, Amazon Web Services and Google Machine Learning Cloud Platform, which attract more developers and promote the vigorous development of machine learning [9].

3. Human-computer Interaction System based on Machine Learning

3.1. Target detection method

Background subtraction method is also called background subtraction method. Its principle is similar to inter-frame subtraction method. Background subtraction method only analyses the difference between current frame and background image[10]. The gray value of video image pixels changes little as background image, while the gray value changes greatly as moving object or target makes movement. Background subtraction method is simple to implement and has low complexity. Its detection effect depends on background image extraction. Because of noise interference such as illumination change and camera jitter, background image may have dynamic changes. Generally, it is necessary to estimate and reconstruct background image by integrating video image sequence information.

3.2. Joint action feature extraction

Human motion feature is the representation of human structure, appearance, background and other objects. It is the basis of classifying and judging action classifier. The feature representation includes many forms, such as whole, local, optical flow and trajectory. Because the experimental background of this paper is simple and the overall outline of human body is clear, the whole geometric feature is adopted. The eight global relative value features of aspect ratio, rectangular ratio, area to convex surface ratio, circumference to convex circumference ratio, spherical characteristic ratio, inner-outer circular ratio, eccentricity and shape parameter are all global geometric features. There are more effective information about human body contour, while Hu invariant moment can determine the center of mass of the image and get no displacement for the image. Variable center moments can satisfy the invariance of translation, scaling and rotation of the target image. The combined features of these two features have their own advantages, which can express human motion well, and make the experiment in this paper can recognize motion quickly and accurately. Generally speaking, the relative value features and Hu invariant moments are not of the same order of magnitude. In order to follow-up data processing, this paper makes a simple normalization of the two.

3.3. Action recognition and simulation

The platform of action recognition algorithm is placed on the computer. Video information of human target action is acquired by the robot's “eyes”. It is transmitted to the computer for analysis and recognition, relevant information is extracted, and the angle arc of the robot is calculated. Finally, it is sent back to the NAO robot for action simulation.

Because there are many human action samples and classification features selected in this paper, a fast and reasonable SVM classifier is chosen to model and classify human action in the next step.

SVM is a two-class supervised classification model. In fact, the classification process is to solve the optimal classification surface. It was originally applied to linear separable cases, but now it has been extended to non-linear cases, and has strong generalization ability.

In this paper, the whole relative geometric moments and Hu invariant moments are used as features to train and learn the learning samples. Finally, SVM is used to classify the static postures, and XY axis dynamic information is used to analyze the human motion to realize the dynamic-static combination of motion recognition method. In the part of target detection and tracking, we get the connected area, human contour information and spatial location information of each frame image target, extract joint features, recognize static posture by multi-class SVM classification, send the results and spatial location information into the state list together, refresh the state, and get the final action description.

The state function sets the sliding window size to 100 frames. When the new frame is included in the state list, the earliest frame in time sequence will be replaced. If the new data frame is incorrect, the list will not be updated. In this way, the recent 100 frame records will always be kept in the state list, and the comprehensive action judgment will be made according to the action recognition results of the 100 frames and the target XY plane information. Discrimination and analysis.

3.4. Robot imitation

NAO robot is a humanoid robot developed by Aldebaran Robotics, France. It is compact and cute in shape, and has a certain level of artificial intelligence. It is the most widely used humanoid robot in the academic field in the world. NAO robots are equipped with a variety of types and numbers of sensors, especially four pressure sensors on the sole of the foot, which can assist in adjusting their movements, making the body balance better and ensuring the smooth movement. In addition to excellent hardware configuration, another major feature of NAO robots is the diversity and comprehensiveness of embedded software, based on NAO robots'own instruction blocks. Users can realize visual programming through the Graphical Programming Software of Clegraphe developed by Aldebaran Robotics (Fig. 5.4), which is common in Windows, Linux and MacOS operating systems. NAO robot also has abundant API interface. Users can use C++, Python language, and can also be controlled by webots, MATLAB and other simulation software. Complete hardware configuration and various software settings enable NAO robots to achieve a variety of functions, such as sound synthesis, positioning, visual image processing, motion control, etc. It has been widely used in the academic field and has become an ideal experimental platform.

3.5. Experimental results and analysis

In this paper, eight global relative geometric moments and seven Hu invariant moments are used as features to train and learn the learning samples. Finally, SVM is used to classify the static postures, and XY axis dynamic information is used to synthetically judge human movements. Firstly, the waving, standing and sitting samples are learned, and the samples are automatically intercepted from the experimental video. After the sample learning is completed, the standing, waving and sitting actions are identified by SVM classifier, and the walking is judged by X-Y axis coordinates.

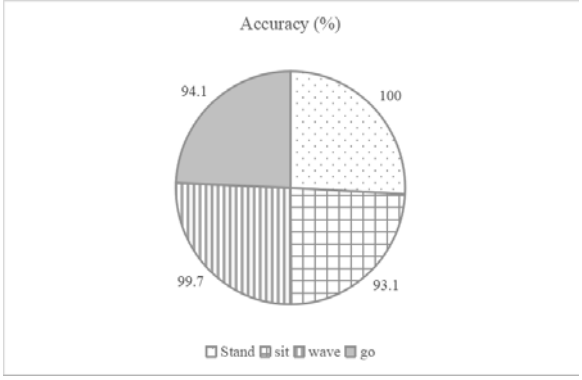

Fig.1. Accuracy of motion recognition

The recognition rate of standing, sitting, waving and walking is shown in Fig. 1. The recognition effect of standing and waving is remarkable, and the recognition of sitting and walking is not susceptible. This is because in training samples, there is no subdivision between sitting and squatting, as long as 15 joint features of human targets are detected. For walking, eamshift tracking logic only traces XY plane displacement information, without detecting angle offset, and state function can not count Z axis displacement information. Once moving to the video capture source, it can not judge walking action, which affects the recognition accuracy.

Table 1 Radius of some joint points imitated by wave action

	RshouldRoll	Lshoulderoll	RshouldPitch	LshouldePitch
Actual radian	-0.212	0.273	1.154	1.280
Imitation radian	-0.458	0.458	1.197	1.196
Error	0.254	0.184	0.042	0.084

The recognition results are sent to the NAO robot for its simulation. The coordinate transformation results and recognition instructions are processed by the control module. The radian information of the joint points is calculated according to the overall relative value characteristics. The NAO robot is imported to simulate the wave motion with 10 frames interval. As shown in Table 1, the actual radian is the real joint radian data obtained by the getAngles interface in the ALMOTION module of NAO robot. The simulated radian is the joint radian data obtained by feature processing calculation and coordinate transformation. It can be seen that the real joint radian and the robot are transmitted. There are errors between simulated joint radians, because when NAO robot completes an action, it actually needs the participation and support of multiple joint degrees of freedom. In addition, the inertial sensor of NAO robot also participates in the fuselage control in real time, while the experimental data calculated are only part of the key node information, which is not fused. It senses data.

4. Conclusion

Humanoid robots have the corresponding “vision”, “hearing”, “touch” and “emotion” with human beings. They can interact with human beings and cooperate with human beings. It is of far-reaching significance to study how to make humanoid robots interact with human beings efficiently and serve human beings. This paper elaborates the frame flow of the combination of static and dynamic motion recognition, realizes the action recognition algorithm, and summarizes and analyses the action recognition rate. In addition, the structure and motion, vision and control module of NAO robot are introduced. The results of action recognition, XY plane information and radian information are imported into the motion control module of NAO robot to simulate the four human actions of standing, sitting, waving and walking.

Acknowledgement

In this paper, the research was sponsored by the Xiamen Overseas Scholar Project (No. XRS2016-314).

References

[1] Wu W L, He W, Feng B. Study on Technology of Triangular Polyhedron 3D Human-Computer Interaction Modeling. Advanced Materials Research, 2014, 962-965:2721-2725.

[2] Nandakumar K, Funk J L. Understanding the timing of economic feasibility: The case of input interfaces for human-computer interaction. Technology in Society, 2015, 43:33-49.

[3] Nobarany S, Booth K S, Hsieh G. What motivates people to review articles? The case of the human-computer interaction community. Journal of the Association for Information Science & Technology, 2015, 67(6):1358-1371.

[4] Nickel M, Murphy K, Tresp V. A Review of Relational Machine Learning for Knowledge Graphs. Proceedings of the IEEE, 2015, 104(1):11-33.

[5] Jean N, Burke M, Xie M. Combining satellite imagery and machine learning to predict poverty. Science, 2016, 353(6301):790.

[6] Malhotra R. A systematic review of machine learning techniques for software fault prediction. Applied Soft Computing Journal, 2015, 27(C):504-518.

[7] Kächele M, Schels M, Meudt S. On Annotation and Evaluation of Multi-modal Corpora in Affective Human-Computer Interaction. Journal of Sexual Medicine, 2014, 8757(7):35-44.

[8] Arruti A, Cearreta I, Alvarez A. Feature selection for speech emotion recognition in Spanish and Basque: on the use of machine learning to improve human-computer interaction.. Plos One, 2014, 9(10):e108975.

[9] Herrera G L, López M, Guerrero L A. Human-Objects Interaction: A Framework for Designing, Developing and Evaluating Augmented Objects.. International Journal of Human-Computer Interaction, 2014, 30(10):787-801.

[10] Amendola S, Bianchi L, Marrocco G. Movement Detection of Human Body Segments: Passive radio-frequency identification and machine-learning technologies. IEEE Antennas & Propagation Magazine, 2015, 57(3):23-37.