MULTI MODAL ROBOT TRAINING USING CGI VIDEOS

Background

Robots require a control program to operate in different environments. The task of coding a control program can be complicated and requires extensive testing in a real time setup leading to damage to the Robot, injuries and complicated setup procedures for every product changeover. The time consumed and the delays associated with manual coding can affect productivity and lead to customer dissatisfaction. Customers require a solution to address the need for fast and reliable deployment of automated systems or Robots to deliver quality products to their customers quickly to maintain competitive edge.

Summary

The following embodiments and aspects thereof are described and illustrated in conjunction with systems and methods which are meant to be exemplary and illustrative, not limiting in scope.

The present disclosure generally relates to Robots utilised in a manufacturing or testing setup that eliminates the need for frequent training, calibration and setup especially when frequent product changes occur in an industrial manufacturing setup. Using CGI (Computer-generated imagery) complemented by CAD (Computer-aided design) process, the control program generates the sequences to operate the Robot to perform complex tasks to process a given object or product, eliminating the need for human intervention. The time consumed in training, configuring and setting up a Robot is significantly reduced as they are built into the CGI and

CAD input to the computer, also referred to as VA (Video & Audio) input, to generate the control program. The ability of a Robot to learn is an important aspect of intelligence, as an automated system without this capability generally cannot be easily deployed in an industrial environment. The present invention disclosed aims to address this need.

The present invention provides a system and methods that may provide for deciphering the operating sequence from the VA input that may include information related to spatial configuration of the processing workspace layout, operational speed and other motion related parameters, location of the source of the objects and its destination, including any intermediate processing location (if any), handling techniques that may include grasping, lifting, orientation and angular movements along with their associated XY&Z coordinates, involved during the movement of the object and a host of image processing related functions to support the Robot for executing the task efficiently. The VA input mimics the virtual representation of the real world processes, through which the computer deciphers the task to be performed and subsequently applies Generative Artificial intelligence techniques and CAD processes by converting the VA input into multiple video frames under their respective task domains as the first step towards generating a control program. A coordinated and synchronous task process is generated for each and every task and combined together to produce a control program that reflects the real world functions of the automated system that closely matches the virtual representation.

Product specific CGI videos are used to build and integrate production architecture, infrastructure and workspace requirements to operate the automated system. Stored CGI videos also represent a knowledge base of pre-programmed recipe files accumulated in a database for ease of scalability and flexibility. Reusability of recipe files aids in quicker configuration and

setup of the system resulting in faster time to market. Incorporating machine learning tools with reinforced learning modules improves the performance of the CGI analysis process for an error free control program associated with a recipe file comprising the relevant parameters that applies to the object or product. The control program further fine tunes itself through the feedback loop (positive or negative), which enhances the performance of the automated system by incorporating the appropriate modifications that affect quality, speed, process control or productivity in real world operation. The effect of the resulting control program ensures a consistent, accurate and reliable operation of the automated system closely matching the CGI's video's virtual representation. Audio specific information embedded within the CGI input is processed by NLP (Natural Language Processor) and incorporates them into the control program in their respective time domains. Audio output at the specified time is automatically processed at the specified moments to gain critical attention, when required.

Brief Description of Drawings

Throughout the drawings, reference numbers may be re-used to indicate correspondence between referenced elements. The drawings are provided to illustrate example embodiments described herein and are not intended to limit the scope of the disclosure.

FIG 1 is an illustration of a articulated Robot in accordance with an embodiment of the present invention;

FIG 2 illustrates a articulated Robot deployed in a automated workspace station in accordance with an embodiment of the present invention;

FIG 3 illustrates the contents of a CGI video, are classified under multiple stages of the automated process extracted in 2D and 3D three-dimensional formats under separate entities, allowing for more realistic representations during the operation of the Robot;

FIG 4 illustrates an example of how the different entities extracted from the CGI input are utilised to generate the control program for the Robot in conjunction with CAD processes before it is stored in computer memory;

FIG 5 illustrates a view of a Robot with its control program and the knowledge base stored in memory for quick reference and reuse when required.

FIG 6 illustrates an isometric view of a Robot with multiple types of Robot hands or end effectors mounted near the Robot for adapting to different types of products;

FIG. 7 is a flow diagram showing the generation of a control program after extracting the entities from the CGI video input file in association with CAD process information.

Detailed Description of Drawings

The present disclosure is generally directed to using CGI video input to a computer in association with a CAD process to generate a control program by extracting different entities and analysing them to perform a set of tasks in an articulated Robot and an integrated automated system, where applicable. The control program that dictates the sequence of actions of the Robot is further able to generate and/or recalibrate itself to enhance the speed, quality and efficiency of a given task without requiring human oversight.

Extracting video frames from the CGI input and classifying them under different entities is the fundamental step in the process, to generate a control program. The entities are grouped under the spatial configuration entity, the motion tracking and operation speed recognition entity, object locations entity, process executed at the specified location entity, the image processing actions entity, the audio analysis entity, workspace layout entity and supporting automation entity. The Robot operation is more accurately extracted, understood and analysed within a CGI input in association with the knowledge deciphered from CAD processes. Essentially CGI video input serves as a virtual representation of the real world processes. The generated control program is fine-tuned using generative AI techniques, in collaboration with machine learning tools aided by reinforcement learning tools, resulting in a reliable and efficient control program for the Robot and supporting automation.

The articulated Robot **100** in **Fig 1** is a typical Robot implemented in automated manufacturing, testing and packaging applications in many types of industries. The system comprises a Camera **10** in combination with illumination module **20** triggered by a illumination controller (not shown) and a pair of hands **30** & **40** that are able to move in X,Y&Z directions with end effectors **50** and **60** mounted at each end of the hands. The end effectors are interchangeable, flexible and scalable to adapt to different types of products to be processed.

In one embodiment of the present invention, **Fig 2** shows a Robot **100** deployed within an automated workspace comprising an input module **85** loaded with objects **80.** The Robot right hand **50** has placed an object **80** at an intermediate station **120** after picking it up from yet another intermediate station **90.** The Robot left hand **60** is in the process of positioning the output station **110** which is loaded with an object that was already processed earlier in station

120. The Robot left end effector **60** subsequently moves to pick up the object **80** placed in the intermediate station **120** after completion of the task specified. The task specified can be for example, inspection for defects, packing the object..etc. The process flow for handling the object **80** is illustrated by **130a**, **130b** and **130c**. The CGI video is created using multi-modal technologies by combining multiple types of data including text, video, audio, and images, to create a comprehensive understanding of the task to be performed in an automated system. This video subsequently becomes the basic building block for a multimodal learning model which reflects a virtual representation of the operation.

In an aspect of the present invention, a method to create a control program based on the CGI video input is disclosed. In **Fig 3**, the method includes the extraction of multiple frames from the CGI input. The CGI input 150 is the input to a computer that comprises the entire operation of the automated system to process a certain object. The computer extracts and classifies the frames starting with positional data mapping in 3D coordinates and motion tracking entity in **160**. The positional data is recorded in each and every frame to determine the spatial configuration of the automated workspace consisting of the input, processing and output station in one embodiment. Furthermore, the positional data is mapped and consolidated using artificial intelligence to plot the trajectory of all movements in 3D coordinates and related sequence timing during the process flow in **160**. The next entity **170** consists of the positions of the object being processed that accurately identifies its location in 3D coordinates at the input, process and output station and any other intermediate position.

The imaging activities entity **180** identifies activities such as quality inspection process, object orientation and manipulation, electrical testing, if the object is an electronic item, biometric identification**,** gesture recognition and any other operation extracted from the CGI input. In

180, activities related to visual display for interactive dashboards that display machine status, performance metrics, and operational data are extracted. In **180,** the process also involves designing intuitive CGI-based interfaces for controlling machines, incorporating visual elements like buttons, sliders, and gauges to facilitate user interaction, if specified. All real time controlling events such as illumination triggers, Camera strobing related to the inspection system are consolidated to form part of the control program. The entity **180** comprises parameters to inspect the object based on identified and specified parameters may also be in the form of embedded text or audio instructions apart from dimensional and pattern details.

Natural language data extracted from the audio stream in the audio analysis entity **190** provides further details of a certain operation that may enhance the functional aspects of all entities by applying the instructions effectively. Audio analytics may further enhance the machine behaviour by implementing visual alerts and notifications when machine parameters deviate from expected norms, allowing for quick intervention. Through the implementation of CALO (Cognitive Assistant that Learns and organises), the audio based commands are incorporated into the control program to execute the necessary task.

The relationships established between each of the entities (**150 to 190**) after applying AI and ML tools, are identified to efficiently create a synchronised control program to perform the task in the automated system that reflects the real world operation as contained in the CGI input.

Fig 4, illustrates an enhanced method of creating a functional control program and storing it in a database. The control program **200** created in **Fig 3** is further fine tuned and corrected for missing or positional data with errors, trajectory pathways to avoid collisions, object

parameters related to shape, size and any other dimensions using CAD process in **210**. The final control program **220** ensures optimum performance of the automated system that has taken into account all aspects of the process resulting in a consistent and reliable operation.

In **Fig 5**, the control program **220** is deployed in the Robot memory (brain) enabling the operation, testing and validation of a physical prototype. In some instances, critical testing and validation is conducted on digital replicas of physical machines that are virtually created which can be monitored and controlled remotely, enabling predictive maintenance and performance optimization. After certification of the control program, the Robot loaded with the certified control program is integrated into the physical automated system.

In **Fig 6**, the Robot **100** is illustrated to demonstrate the scalable and flexible aspects. The Robot **100** comprises different types of end effectors **260** and **300** mounted on end effector stands **240.** End effectors are designed to be easily mounted at position **280** of the Robot hand and can be quickly interchanged to handle different objects. The Robot **100** is programmed to safely engage and disengage with the end effectors as and when required.

Fig 7 shows a schematic diagram **300** of an illustrative flow to generate a control program beginning with step **320** followed by the CGI input **325** to a computer **330.** The computer **330** subsequently begins the process of extracting different entities from the CGI input starting with Spatial configuration entity, comprising all the motion trajectories **336**, the entity **340** with details of object locations in the workspace, the imaging activities entity **344** comprising quality control processes, the entity **334** related to the supporting automation and their role in the process flow, the entity **338** with the data related to layout of the workspace along with integrated accessories and entity **342** containing an audio stream embedded with audio and

other behavioural instructions to be incorporated into the control program. In step **348**, the information within all entities are further processed using computer aided design (CAD) processes to fine tune object related parameters such as positional data, trajectory pathways, object parameters related to shape, size and any other dimensions. The generated control program subsequently uploaded to the Robot controller in step **350** to operate the Robot integrated with the automated system. During the operation of the Robot in **352** and its associated supporting automated system, the control program further regulates and corrects its operation by using machine learning and reinforced learning tools in **354** and **358** respectively. In **360** the Generative AI visualises the operation of the machine, identifies bottlenecks, and optimises the workflow by implementing corrections and modifications to generate a new version of the control program in **362** and subsequently upload the revised control program to the Robot controller. This process of continuous learning through the use of convolutional neural networks (CNN) applied to text, audio and video is utilised to update the process flow. The computer further evaluates the revised control program by monitoring the movements of the automated system visually along with the sensor data and provides feedback relating to the success of the update. Subsequently, each successful update is assigned an award point depending upon the complexity and its efficiency in resolving an issue or improving the performance of any of the entities. The rewarding system operates as part of steps **354** and **358** by suggesting updates to the control program in step **360** which in turn decides if the updates or improvements are worth implementing by evaluating their respective reward points. Furthermore, the control program is incorporated with predictive analytics techniques to enable it to identify the likelihood of future outcomes based on accumulated historical data. Predictive maintenance, energy consumption, throughput limitations and safety aspects are some of the issues highlighted by predictive analytics techniques. The finalised updates may enhance any

number of performance parameters such as speed, quality, accuracy, consistency, sustainability and efficiency to be eligible for generation of a new version of the control program in step **362.**

References in this description to “an embodiment,” “one embodiment,” or the like mean that a particular feature, structure, material, or characteristic being described is included in at least one embodiment of the present invention. Thus, the appearances of such phrases in this specification do not necessarily all refer to the same embodiment. On the other hand, such references are not necessarily mutually exclusive either. Furthermore, the particular features, structures, materials, or characteristics can be combined in any suitable manner in one or more embodiments. It is to be understood that the various features shown in the figures are merely illustrative representations.

Claims:

1. A control program generation method for an automated system using a computer, the method comprising:

extracting video frames using a computer generated imagery (CGI) video as the input source;

the video frames further classified and grouped under multiple entities such as spatial configuration entity, trajectory tracking and operation speed recognition entity, object locations entity, process executed at the specified location entity, image activities entity, audio analysis entity, workspace layout entity and supporting automation entity;

a computer aided design (CAD) process to complement the generation of the control program;

analysing the spatial configuration entity to determine the positional data in 3D coordinates of the input, processing and output station;

analysing the trajectory tracking and operation speed recognition entity to determine the movement in 3D coordinates and sequence timing during the process flow;

analysing the object location entity, to identify and map all the locations such as input, process and output station and any other intermediate positions;

analysing the imaging activities entity, to execute the given task as identified;

analysing the imaging activities entity, to design an intuitive CGI-based interface for controlling the automated system by incorporating visual elements like buttons, sliders, and gauges to facilitate user interaction.

analysing the imaging activities entity to control the vision system trigger controller to strobe the illumination and camera shutter for image capture to inspect the quality of the object, based on identifiable parameters which may be in the form of embedded text or audio instructions.

analysing the audio analysis entity, to extract natural language data from the audio stream that provides details of a certain operation such as audio commands, instructions or alerting sounds to enhance the functional aspects;

incorporation of CALO (Cognitive Assistant that Learns and organises) techniques to enable the control program understand audio related commands clearly to produce an effective response;

creation of replicas of the automated system as contained in the CGI video that can be monitored and controlled remotely, enabling predictive maintenance and performance optimization;

incorporating predictive analytics techniques to enable the control program to identify the likelihood of future outcomes based on accumulated historical data;

Generating a control program that accurately models the behaviour of the automated system that allows for testing and validation without a physical prototype;

Generating a control program that reflects the real world function of the automated system as specified in the CGI video that closely matches the virtual representation.

2. The method of claim 1, further comprising:

using Generative artificial intelligence (Gen-AI) combined with machine learning and reinforcement learning tools to visualise the operation of the machine and identify bottlenecks, to optimise the workflow by implementing corrections and modifications resulting in a revised version of the control program and associated recipe file comprising the relevant parameters for the specific product.

3. The method of claim 1, further comprising:

Enhancing the efficiency of the control program through the use of convolutional neural networks (CNN) applied to text, audio and video to update the process flow.

4. The method of claim 1, further comprising:

every successful update awarded with a reward point depending upon the complexity and its efficiency in resolving an issue or improving the performance in any of the entities.

5. The method of claim 1, further comprising:

a process of continuous learning of the process flow achieved through a rewarding system whereby GenAI decides if the program update is worth implementing by way of reward points awarded to every successful update.

6. The method of claim 1, further comprising:

the finalised set of updates to the control program that may enhance any number of performance parameters such as speed, quality, accuracy, consistency, sustainability and efficiency.

7. The method of claim 1, further comprising:

analysis of the object location entity with the aid of computer aided design processes to fine tune object related parameters such as positional data, trajectory pathways, object parameters related to shape, size and any other dimensions.

8. The method of claim 1, further comprising:

multiple CGI videos representing a knowledge base of pre-programmed recipe files accumulated in a database for ease of scalability and flexibility.

9. The method of claim 1, further comprising:

recipe files enabling reusability, quicker configuration and setup of the automated system resulting in faster time to market.

10. An automated system comprising:

an articulated Robot consisting of at least two hands configured to perform a task according to a control program for a specific product;

a computer programmed to execute the control program for the automated system to perform the task of moving the object, processing the inspection of the object, controlling the vision system to perform the assigned task;

a computer to extract and classify a CGI input into different entities to generate a control program to operate a real world automated system;

a dashboard to display the parameters and all other real time data for monitoring;

an imaging system integrated to a trigger controller (not shown) to strobe the illumination module and a camera for capture images;

a pair of Robot hands with the ability to move in X,Y & Z directions;

a pair of end effectors at each end of the Robot hand that are interchangeable and scalable to adapt to different types of objects;

multiple end effector stands holding different types of end effectors located within the reach of the Robot to aid in easy exchange during product changeovers.

11. The automated system of claim 10, wherein the automated system executing the control program:

complements the role played by the convolution neural networks (CNN) in correcting errors of the control program by visually monitoring the movements and sensor data of the automated system and providing feedback to the control program related to success of the updates.

12. The automated system of claim 10, wherein the automated system executing the control program:

incorporating predictive analytics techniques to enable the control program to identify the likelihood of future outcomes based on accumulated historical data.

13. The automated system of claim 12, wherein the automated system executing the control program:

predicts issues related to maintenance, energy consumption, throughput limitations and safety aspects.