(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
07 March 2019 (07.03.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/045779 A1

(51) International Patent Classification:
G06F 15/18 (2006.01)

(21) International Application Number:
PCT/US2018/026122

(22) International Filing Date:
04 April 2018 (04.04.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/553,589 01 September 2017 (01.09.2017) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: **GOLDBERG, Kenneth, Yigael**; 217 Sycamore Avenue, Mill Valley, CA 94941 (US). **MAHLER, Jeffrey, Brian**; 2300 Ward Street, Berkeley, CA 94705 (US). **MATL, Matthew**; 1423 Walnut Street, Apt. 3, Berkeley, CA 94709 (US).

(74) Agent: **WAGNER, Justin, D.** et al.; Marger Johnson, 888 SW 5th Avenue, Suite 1050, Portland, OR 97204 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

(54) Title: ROBOTIC SYSTEMS AND METHODS FOR ROBUSTLY GRASPING AND TARGETING OBJECTS

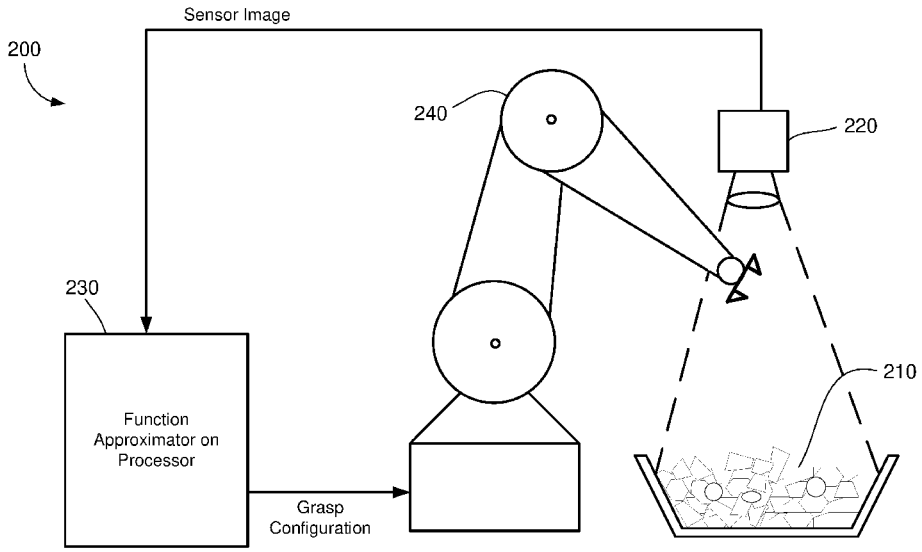

FIG. 2

(57) Abstract: Embodiments are generally directed to generating a training dataset of labelled examples of sensor images and grasp configurations using a set of three-dimensional (3D) models of objects, one or more analytic mechanical representations of either or both of grasp forces and grasp torques, and statistical sampling to model uncertainty in either or both sensing and control. Embodiments can also include using the training dataset to train a function approximator that takes as input a sensor image and returns data that is used to select grasp configurations for a robot grasping or targeting mechanism.

[Continued on next page]

WO 2019/045779 A1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

ROBOTIC SYSTEMS AND METHODS FOR ROBUSTLY GRASPING AND TARGETING OBJECTS

Cross-Reference to Related Application

[0001] This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/553,589, titled "ROBOTIC SYSTEM FOR ROBUSTLY GRASPING OBJECTS" and filed on September 1, 2017, the content of which is hereby fully incorporated by reference herein.

Technical Field

[0002] The disclosed technology relates generally to robotic systems configured for robustly grasping and/or targeting a variety of objects, e.g., in warehouses or homes, e.g., using grippers or suction cup devices and, more particularly, to using libraries of 3D object models that can be analyzed and statistically sampled to train a robot to grasp and/or target objects with robustness to errors in sensing and control.

Background

[0003] Picking objects up is such a fundamental skill for robots that it is sometimes difficult to understand how challenging grasping still is. Robots in factories depend on high quality sensor data along with some amount of advance knowledge about the objects that they will grasp. However, it is much more challenging to design a system that can reliably pick up a variety of previously unseen objects, including the infinitely long tail of objects that can be, for any of a number of reasons, difficult to grasp.

[0004] One attempt around such problems is to design specialized grasping hardware (such as enveloping grasps or adhesives, for example) to compensate for not completely knowing the best way to pick up a given object, but this limits visibility of the object in the gripper.

[0005] Grasping can be simplified when one has an exact model of the object to be grasped and the exact position and location of the object gripper and a gripper that works exactly as expected to work. Sensors are often inaccurate and noisy, and grippers themselves generally have finite amounts of accuracy and precision with which they can be controlled. As such, there is enough uncertainty that consistent robust grasping is a significant, if not unattainable, challenge.

[0006] One approach is to use machine learning. One approach is to train a system to predict how robust a particular grasp on a given object will be (e.g., whether the grasp will fail when the object is lifted or moved) using datasets collected from millions of robots grasping millions of objects in millions of physical robot trials. Unfortunately, this is simply not a practical solution.

Summary

[0007] Implementations of the disclosed technology can be advantageously used in a variety of industries, such as in connection with robots used for warehouse order fulfillment (e.g., where orders are unique and there can be millions of different products), manufacturing, packing, inspection, marking or otherwise labeling, bin picking, and in homes for menial tasks such as decluttering, for example.

Brief Description of the Drawings

[0008] FIGURE 1 illustrates an example of a computer-implemented method in accordance with certain embodiments of the disclosed technology.

[0009] FIGURE 2 illustrates an example of a system in accordance with certain embodiments of the disclosed technology.

Detailed Description

[0010] FIGURE 1 illustrates an example of a computer-implemented method 100 in accordance with certain embodiments of the disclosed technology. In the example, the method 100 includes generating a training dataset 140 of labelled examples 150 and 160 of sensor images and grasp configurations using a set of three-dimensional (3D) models of objects 110, one or more analytic mechanical representations of either or both of grasp forces and grasp torques, and statistical sampling to model uncertainty in either or both sensing and control.

[0011] The method 100 can also include using the training dataset 140 to train a function approximator 170 that takes as input a sensor image 130 and returns data that is used to select grasp configurations 120 for a robot grasping mechanism. In certain embodiments, grasp configurations may be defined by one or more points relative to the sensor image.

[0012] In certain embodiments, the generated training dataset 140 may include a measure of grasp quality associated with a grasp configuration 120. In such embodiments, the measure of grasp quality may include robustness to errors in position or forces. Alternatively or in addition thereto, the measure of grasp quality may include probability of successful lifting of the object.

[0013] In certain embodiments, the statistical sampling may include uncertainty in variables related to initial state, contact, physical motion, friction, inertia, object shape, robot control, sensor data, or any suitable combination thereof.

[0014] In certain embodiments, the function approximator may be a Convolutional Neural Network (CNN), a Random Forest, a Support Vector Machine (SVM), or a linear weight matrix.

[0015] In certain embodiments, the robotic grasping mechanism can include a robot gripper, a multi-fingered robot hand, one or more suction cups, a magnet, adhesive material, or any suitable combination thereof.

[0016] In certain embodiments, the set of 3D object models can include mass properties for computation of stable poses for each object. Alternatively or in addition thereto, the set of 3D object models can include mass properties for computation of resistance to either or both of gravity and inertia for a given grasp and object combination. Alternatively or in addition thereto, the set of 3D object models can include material properties for computation including frictional properties, deformation, porosity, color textures for photorealistic rendering, or any combination thereof.

[0017] In certain embodiments, the set of 3D object models may include "adversarial" objects whose geometry makes them difficult to be grasped.

[0018] In certain embodiments, the set of 3D object models may be augmented by synthetic product packaging models, such as "skin packs" or "blister packs," for example. Alternatively or in addition thereto, the set of 3D object models may be augmented using transformations of the initial set of object models using any of the following operations: scaling, stretching, twisting, shearing, cutting, and combining objects.

[0019] In certain embodiments, the analytic mechanical representations may include wrench mechanics. Alternatively or in addition thereto, the analytic mechanical representations may include any of the following metrics: force closure grasp quality, Ferrari-Canny grasp quality, suction cup contact quality, wrench resistance quality, magnetic contact quality, and adhesive contact quality.

[0020] Such implementations may include using computer simulation to estimate grasp quality. Alternatively or in addition thereto, such implementations may include using statistical sampling to compute statistics of a grasp quality metric. In such embodiments, the grasp quality metric may include average, median, moments, percentiles, or any suitable combination thereof. Alternatively or in addition thereto, such implementations may include using statistical sampling to compute the statistics of the grasp quality metric includes using numerical integration.

[0021] Certain implementations may include using statistical sampling to generate variations in possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques, robot hardware parameters to generate the sensor images (e.g., a 3D depth map), or any suitable combination thereof.

[0022] Certain implementations may include using computer simulation to generate variations in possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques, and robot hardware parameters to generate the sensor images (e.g., a 3D depth map), or any suitable combination thereof.

[0023] In certain embodiments, the set of object models and the computed grasp configurations for each object model may be stored as a network with one or more computed relationships between object models such as similarity in shape. In such embodiments, the data in the network of object models and the computed grasp configurations for each object model may be used to efficiently compute one or more desired grasp configurations for one or more new object models.

[0024] Certain implementations may include communicating over a network, e.g., the Internet, to obtain updated function approximator parameters.

[0025] Certain implementations may include updating the parameters of the function approximator based on outcomes of physical grasp attempts.

[0026] In certain embodiments, a computer-implemented method may include generating a training dataset of labelled examples of sensor images and target points within those images using a set of three-dimensional (3D) models of objects, one or more analytic evaluation methods of desired target points on an object, and statistical sampling to model uncertainty in either or both sensing and control. The method may also include using the training dataset to train a function approximator that takes as input a sensor image and returns data to compute one or more target points for a robot targeting mechanism (e.g., placing a label on the object, affixing a stamp to the object, and inspecting the object).

[0027] In certain embodiments, the statistical sampling may include uncertainty in variables related to initial state, contact, physical motion, friction, inertia, object shape, robot control, sensor data, or a combination thereof.

[0028] The function approximator may include a Convolutional Neural Network (CNN), a Random Forest, a Support Vector Machine, or a linear weight matrix.

[0029] In certain embodiments, the set of 3D object models can include mass properties for computation of stable poses for each object. Alternatively or in addition thereto, the set of 3D object models can include mass properties for computation of resistance to either or both of gravity and inertia for a given targeting and object combination. Alternatively or in addition thereto, the set of 3D object models can include material properties for computation including frictional properties, deformation, porosity, color textures for photorealistic rendering, or any combination thereof.

[0030] In certain embodiments, the set of 3D object models may include "adversarial" objects whose geometry makes them difficult to be targeted.

[0031] In certain embodiments, the set of 3D object models may be augmented by synthetic product packaging models, such as "skin packs" or "blister packs," for example. Alternatively or in addition thereto, the set of 3D object models may be augmented using

transformations of the initial set of object models using any of the following operations: scaling, stretching, twisting, shearing, cutting, and combining objects.

[0032] In certain implementations, the analytic mechanical representations may include wrench mechanics.

[0033] Certain implementations may include using statistical sampling to generate variations in possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques, robot hardware parameters to generate the sensor images (e.g., a 3D depth map), or any suitable combination thereof.

[0034] Certain implementations may include using computer simulation to generate variations in possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques, and robot hardware parameters to generate the sensor images (e.g., a 3D depth map), or any suitable combination thereof.

[0035] In certain embodiments, the set of object models and the computed grasp configurations for each object model may be stored as a network with one or more computed relationships between object models such as similarity in shape. In such embodiments, the data in the network of object models and the computed grasp configurations for each object model may be used to efficiently compute one or more desired grasp configurations for one or more new object models.

[0036] Certain implementations may include communicating over a network, e.g., the Internet, to obtain updated function approximator parameters.

[0037] In certain embodiments, grasp configurations may be defined by one or more points relative to the sensor image.

[0038] FIGURE 2 illustrates an example of a system 200 for grasping objects 210 in accordance with certain embodiments of the disclosed technology. In the example, the system 200 includes a sensor 220, a robotic grasping mechanism 240, and one or more

processors configured to use sensor images to compute a desired grasp configuration for the robotic grasping mechanism 240 based at least in part on a function approximator 230 that is trained on a training dataset of labelled examples of sensor images and grasp configurations using a set of three-dimensional (3D) models of objects, one or more analytic mechanical representations of either or both of grasp forces and grasp torques, and statistical sampling to model uncertainty in either or both sensing and control. In certain embodiments, the function approximator 230 may be configured to take as input a sensor image and return data that is used to compute robust grasp configurations for the robotic grasping mechanism 240.

[0039] In certain embodiments, a set of candidate grasp configurations may be computed based on computing potential antipodal grasps from the sensor image, e.g., a 2D image from a camera.

[0040] Certain implementations may include obtaining a 3D depth map using a structured lighting system, a Lidar system, a stereo pair of color cameras, a stereo pair of monochrome cameras, a monocular image, or any suitable combination thereof.

[0041] Certain implementations may include the use of motion planning methods to avoid robot contact with the environment based on collision checking. Alternatively or in addition thereto, the system may include means for using robot motions to move grasped objects into specific new configurations. Alternatively or in addition thereto, the system may include means for detecting the outcome of grasp attempts. In such embodiments, the system may be configured to detect the outcome of grasp attempts includes using one or more load cells, a light sensor, a camera, a force sensor, a tactile sensor, or any suitable combination thereof.

[0042] Certain embodiments may include applying multiple grasping methods in parallel.

[0043] Certain implementations may include two or more grasping methods where the outputs of two or more function approximators are combined to select between these grasping methods for a given sensor image.

[0044] In certain alternative implementations, the system may include a sensor, a robotic targeting mechanism, and one or more processors configured to use sensor images to compute a desired target configuration for the robotic targeting mechanism based at least in part on a function approximator that is trained on a training dataset of labelled examples of sensor images and target points using a set of three-dimensional (3D) models of objects, one or more analytic mechanical representations of target points, and statistical sampling to model uncertainty in either or both of sensing and control. In such embodiments, the function approximator may be configured to take as input a sensor image (e.g., a 2D image from a camera) and return data that is used to compute robust target points for the robotic targeting mechanism.

[0045] Certain implementations may include obtaining a 3D depth map using a structured lighting system, a Lidar system, a stereo pair of color cameras, a stereo pair of monochrome cameras, a monocular image, or any suitable combination thereof.

[0046] Certain implementations may include the use of motion planning methods to avoid robot contact with the environment based on collision checking. Alternatively or in addition thereto, the system may include means for detecting the outcome of targeting attempts. In such embodiments, detecting the outcome of targeting attempts may include using one or more load cells, a light sensor, a camera, a force sensor, a tactile sensor, or any suitable combination thereof.

[0047] Certain implementations may include applying multiple targeting methods in parallel.

[0048] Certain implementations may include two or more targeting methods where the outputs of two or more function approximators may be combined to select between these targeting methods for a given sensor image.

[0049] Certain implementations may also include pushing objects to separate them from the environment and create an accessible grasp.

[0050] Grasp and target success can be predicted directly from depth images by training a deep Convolutional Neural Network (CNN) on a massive dataset of parallel-jaw grasps, grasp metrics, and rendered point clouds generated using analytic models of robust grasping and image formation.

[0051] To reduce data collection time for deep learning of robust robotic grasp plans, a deep neural network can be trained for grasp classification, e.g., from a synthetic dataset of over 6.7 million point clouds, grasps, and robust analytic grasp metrics generated from thousands of three-dimensional (3D) models in randomized poses on a table. The resulting dataset can be used to train a Grasp Quality Convolutional Neural Network (GQ-CNN) model that rapidly classifies grasps as robust from depth images and the position, angle, and height of the gripper above a table.

[0052] The Grasp Quality Convolutional Neural Network (GQ-CNN) model can be trained to classify robust grasps in depth images using expected epsilon quality as supervision, where each grasp is specified as a 3D pose and depth relative to a camera. A grasp planning method can sample antipodal grasp candidates and rank them with a GQ-CNN.

[0053] An enormous dataset can be used to provide the data to train a neural network to develop highly reliable robot grasping across a wide variety of rigid objects. For example, the dataset can consist of up to or over 6.7 million point object point clouds, accompanying parallel-jaw gripper poses, along with a robustness estimate of how likely it is that the grasp will be able to lift and carry the object. This can be advantageously used in a robust robotic grasping system.

[0054] Implementations of the disclosed technology can rely on a probabilistic model to generate synthetic point clouds, grasps, and grasp robustness labels from datasets of 3D object meshes using physics-based models of grasping, image rendering, and camera noise, thus leveraging cloud computing to rapidly generate a large training set for a CNN.

[0055] Implementations can include a hybrid approach to machine learning that combines physics with Deep Learning, e.g., by combining a large dataset of 3D object shapes, a physics-based model of grasp mechanics, and sampling statistics to generate many (e.g., 6.7 million) training examples, and then using a Deep Learning network to learn a function that can rapidly find robust grasps when given a 3D sensor point cloud. The system can be trained on a very large set of examples of robust grasps, similar to recent results in computer vision and speech recognition.

[0056] In situations where the CNN estimates the robustness to be high, the grasp generally works as expected. For example, in certain experiments with an ABB YuMi (i.e., a two-arm industrial robot), the planner was 93 percent successful in planning reliable grasps and was also able to successfully predict grasp robustness with 40 novel objects (including tricky things like a can opener and a washcloth) with just one false positive out of 69 predicted successes.

[0057] Since a robot may have a good idea as to when it will succeed, it may also be able to tell when it is likely to fail. In situations where the robot anticipates a failure, the robot could take appropriate action, e.g., by either poking the object to change its orientation or asking a human for help.

[0058] Implementations may be compatible with virtually any 3D camera and parallel-jaw or suction-cup gripper, and may be used to choose a primary grasp axis for multi-fingered grippers.

[0059] In certain implementations, certain knowledge specific to the hardware setup, such as the focal length and bounds on where the RGB-D sensor will be relative to the robot, the geometry of a parallel-jaw robot gripper (specified as CAD model), and a friction coefficient for the gripper, may be provided as input to generate a new training dataset specific to a given

hardware setup. In such implementations, a GQ-CNN trained on the dataset may have successful performance.

[0060] Implementations of the disclosed technology may include using robust analytic grasp metrics as supervision, using the gripper's distance from the camera in predictions, and performing extensive evaluations on a physical robot.

[0061] Certain implementations may facilitate development of new architectures for predicting grasp robustness from point clouds, and also to encourage the benchmarking of new methods.

[0062] Certain implementations may include automatically generating training datasets for robotic grasping that can be a useful resource to train deep neural networks for robot grasp planning across multiple different robots.

[0063] In certain implementations, a robot may be integrated with an artificial intelligence so that it can figure out how to robustly grip objects it has never seen before or otherwise encountered, for example.

[0064] In certain implementations, a robot may use a neural network and a sensor (e.g., a Microsoft Kinect 3D sensor) to see a new object and then determine a robust grasp for successfully grasping the object.

[0065] Certain implementations may include household robots performing various chores such as vacuuming, doing dishes, and picking up clutter, for example. Such machines will frequently encounter new objects but, by teaching themselves, they can better adapt to their surroundings.

[0066] In certain implementations, robots may be communicating via the cloud, e.g., to share information amongst each other, rather than working and learning in isolation. In such implementations, a robot can distribute gained knowledge to other robots that are like it and, in certain embodiments, even entirely different kinds of robots.

[0067] As used herein, the Dexterity Network (Dex-Net) generally refers to a research project that includes code, datasets, and algorithms for generating datasets of synthetic point clouds, robot parallel-jaw grasps, and metrics of grasp robustness based on physics for up to or over thousands of 3D object models to train machine learning-based methods to plan robot grasps. Implementations may include developing highly reliable robot grasping across any of a wide variety of rigid objects such as tools, household items, packaged goods, and industrial parts.

[0068] Dex-Net 1.0 may be used for learning predictors of grasp success for new 3D mesh models, e.g., to accelerate generation of new datasets. Dex-Net 2.0 may be used for learning Grasp Quality Convolutional Neural Network (GQ-CNN) models that predict the probability of success of candidate grasps on objects from point clouds. GQ-CNNs may be useful for quickly planning grasps that can lift and transport a wide variety of objects by a physical robot.

[0069] Certain implementations may include an analytic suction grasp model and metric based on the set of wrenches at the contact interface between the suction cup and rigid object surface and the magnitude of contact wrenches that needed to resist external wrenches due to gravity under perturbations in object pose, center of mass, contact location, friction, and gravity. This metric can be used to generate a dataset of up to or over 6.7 million point clouds, suction grasps, and grasp robustness labels generated from up to or over 1,500 3D object models, and train a Grasp Quality Convolutional Neural Network (GQ-CNN) on this dataset to classify grasp robustness from point clouds.

[0070] Having described and illustrated the principles of the invention with reference to illustrated embodiments, it will be recognized that the illustrated embodiments may be modified in arrangement and detail without departing from such principles, and may be

combined in any desired manner. And although the foregoing discussion has focused on particular embodiments, other configurations are contemplated.

[0071] Consequently, in view of the wide variety of permutations to the embodiments that are described herein, this detailed description and accompanying material is intended to be illustrative only, and should not be taken as limiting the scope of the invention. What is claimed as the invention, therefore, is all such modifications as may come within the scope and spirit of the following claims and equivalents thereto.

What is claimed is:

1. A computer-implemented method, comprising:

generating a training dataset of labelled examples of sensor images and grasp configurations using:

a set of three-dimensional (3D) models of objects;

one or more analytic mechanical representations of either or both of grasp forces and grasp torques; and

statistical sampling to model uncertainty in either or both sensing and control; and

using the training dataset to train a function approximator that takes as input a sensor image and returns data that is used to select grasp configurations for a robot grasping mechanism.

2. The computer-implemented method of claim 1, wherein the generated training dataset includes a measure of grasp quality associated with a grasp configuration.

3. The computer-implemented method of claim 2, wherein the measure of grasp quality includes robustness to errors in position or forces.

4. The computer-implemented method of claim 2, wherein the measure of grasp quality includes probability of successful lifting of the object.

5. The computer-implemented method of claim 1, wherein the statistical sampling includes uncertainty in variables related to at least one selected from a group consisting of: initial state, contact, physical motion, friction, inertia, object shape, robot control, and sensor data.

6. The computer-implemented method of claim 1, wherein the function approximator is selected from a group consisting of: a Convolutional Neural Network (CNN), a Random Forest, a Support Vector Machine, and a linear weight matrix.

7. The computer-implemented method of claim 1, wherein the robotic grasping mechanism includes at least one selected from a group consisting of: a robot gripper, a multi-fingered robot hand, one or more suction cups, a magnet, and adhesive material.

8. The computer-implemented method of claim 1, wherein the set of 3D object models includes mass properties for computation of stable poses for each object.

9. The computer-implemented method of claim 1, wherein the set of 3D object models includes mass properties for computation of resistance to either or both of gravity and inertia for a given grasp and object combination.

10. The computer-implemented method of claim 1, wherein the set of 3D object models includes material properties for computation of at least one of a group consisting of: frictional properties, deformation, porosity, and color textures for photorealistic rendering.

11. The computer-implemented method of claim 1, wherein the set of 3D object models is augmented by synthetic product packaging models.

12. The computer-implemented method of claim 11, wherein the synthetic product packaging models include “skin packs” or “blister packs.”

13. The computer-implemented method of claim 1, wherein the set of 3D object models is augmented using transformations of the initial set of object models using at least one operation selected from a group consisting of: scaling, stretching, twisting, shearing, cutting, and combining objects.

14. The computer-implemented method of claim 1, wherein the set of 3D object models includes “adversarial” objects whose geometry makes them difficult to grasp.

15. The computer-implemented method of claim 1, wherein the analytic mechanical representations include wrench mechanics.

16. The computer-implemented method of claim 1, wherein analytic mechanical representations include at least one metric selected from a group consisting of: force closure grasp quality, Ferrari-Canny grasp quality, suction cup contact quality, wrench resistance quality, magnetic contact quality, and adhesive contact quality.

17. The computer-implemented method of claim 16, further comprising using computer simulation to estimate grasp quality.

18. The computer-implemented method of claim 16, further comprising using statistical sampling to compute statistics of a grasp quality metric.

19. The computer-implemented method of claim 18, wherein the grasp quality metric includes at least one selected from a group consisting of: average, median, moments, and percentiles.

20. The computer-implemented method of claim 18, wherein using statistical sampling to compute the statistics of the grasp quality metric includes using numerical integration.

21. The computer-implemented method of claim 1, further comprising using statistical sampling to generate variations in at least one of a group consisting of: possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques, and robot hardware parameters to generate the sensor images.

22. The computer-implemented method of claim 1, further comprising using computer simulation to generate variations in at least one of: possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques and robot hardware parameters to generate the sensor images.

23. The computer-implemented method of claim 1, wherein the sensor image includes a 3D depth map.

24. The computer-implemented method of claim 1, wherein the set of object models and the computed grasp configurations for each object model is stored as a network with one or more computed relationships between object models such as similarity in shape.

25. The computer-implemented method of claim 24, wherein the data in the network of object models and the computed grasp configurations for each object model is used to efficiently compute one or more desired grasp configurations for one or more new object models.

26. The computer-implemented method of claim 1, further comprising communicating over a network to obtain updated function approximator parameters.

27. The computer-implemented method of claim 26, wherein the network includes the Internet.

28. The computer-implemented method of claim 1, further comprising updating the parameters of the function approximator based on outcomes of physical grasp attempts.

29. The computer-implemented method of claim 1, wherein grasp configurations are defined by one or more points relative to the sensor image.

30. A computer-implemented method, comprising:

generating a training dataset of labelled examples of sensor images and target points within those images using:

a set of three-dimensional (3D) models of objects;

one or more analytic evaluation methods of desired target points on an object; and

statistical sampling to model uncertainty in either or both sensing and control; and

using the training dataset to train a function approximator that takes as input a sensor image and returns data to compute one or more target points for a robot targeting mechanism.

31. The computer-implemented method of claim 30, wherein the robot targeting mechanism includes at least one selected from a group consisting of: placing a label on the object, affixing a stamp to the object, and inspecting the object.

32. The computer-implemented method of claim 30, wherein the statistical sampling includes uncertainty in variables related to at least one selected from a group consisting of: initial state, contact, physical motion, friction, inertia, object shape, robot control, and sensor data.

33. The computer-implemented method of claim 30, wherein the function approximator is selected from a group consisting of: a Convolutional Neural Network (CNN), a Random Forest, a Support Vector Machine, and a linear weight matrix.

34. The computer-implemented method of claim 30, wherein the set of 3D object models includes mass properties for computation of stable poses for each object.

35. The computer-implemented method of claim 30, wherein the set of 3D object models includes mass properties for computation of resistance to either or both of gravity and inertia for a given targeting and object combination.

36. The computer-implemented method of claim 30, wherein the set of 3D object models includes material properties for computation of at least one of a group consisting of: frictional properties, deformation, porosity, and color textures for photorealistic rendering.

37. The computer-implemented method of claim 30, wherein the set of 3D object models is augmented by synthetic product packaging models.

38. The computer-implemented method of claim 37, wherein the synthetic product packaging models include “skin packs” or “blister packs.”

39. The computer-implemented method of claim 30, wherein the set of 3D object models is augmented using transformations of the initial set of object models using at least one operation selected from a group consisting of: scaling, stretching, twisting, shearing, cutting, and combining objects.

40. The computer-implemented method of claim 30, wherein the set of 3D object models includes “adversarial” objects whose geometry makes them difficult to target.

41. The computer-implemented method of claim 30, wherein the analytic mechanical representations include wrench mechanics.

42. The computer-implemented method of claim 30, further comprising using statistical sampling to generate variations in at least one of a group consisting of: possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques, and robot hardware parameters to generate the sensor images.

43. The computer-implemented method of claim 30, further comprising using computer simulation to generate variations in at least one of: possible object poses, sensor poses, camera parameters, lighting, material properties, friction, forces, torques and robot hardware parameters to generate the sensor images.

44. The computer-implemented method of claim 30, wherein the sensor image includes a 3D depth map.

45. The computer-implemented method of claim 30, wherein the set of object models and the computed targeting configurations for each object model is stored as a network with one or more computed relationships between object models such as similarity in shape.

46. The computer-implemented method of claim 45, wherein the data in the network of object models and the computed targeting configurations for each object

model is used to efficiently compute one or more desired targeting configurations for one or more new object models.

47. The computer-implemented method of claim 30, further comprising communicating over a network to obtain updated function approximator parameters.

48. The computer-implemented method of claim 47, wherein the network includes the Internet.

49. The computer-implemented method of claim 30, wherein grasp configurations are defined by one or more points relative to the sensor image.

50. An apparatus, comprising:

a sensor;

a robotic grasping mechanism; and

one or more processors configured to use sensor images to compute a desired grasp configuration for the robotic grasping mechanism based at least in part on a function approximator that is trained on a training dataset of labelled examples of sensor images and grasp configurations using:

a set of three-dimensional (3D) models of objects;

one or more analytic mechanical representations of either or both of grasp forces and grasp torques; and

statistical sampling to model uncertainty in either or both sensing and control;

wherein the function approximator is configured to take as input a sensor image and return data that is used to compute robust grasp configurations for the robotic grasping mechanism.

51. The apparatus of claim 50, wherein a set of candidate grasp configurations is computed based on computing potential antipodal grasps from the sensor image.

52. The apparatus of claim 50, further comprising obtaining a 3D depth map using at least one selected from a group consisting of: a structured lighting system, a Lidar system, a stereo pair of color cameras, a stereo pair of monochrome cameras, and a monocular image.

53. The apparatus of claim 50, wherein the sensor image includes a 2D image from a camera.

54. The apparatus of claim 50, further comprising use of motion planning methods to avoid robot contact with the environment based on collision checking.

55. The apparatus of claim 50, further comprising means for using robot motions to move grasped objects into specific new configurations.

56. The apparatus of claim 50, further comprising means for detecting the outcome of grasp attempts.

57. The apparatus of claim 56, wherein detecting the outcome of grasp attempts includes using at least one selected from a group consisting of: one or more load cells, a light sensor, a camera, a force sensor, and a tactile sensor.

58. The apparatus of claim 50, further comprising applying multiple grasping methods in parallel.

59. The apparatus of claim 50, further comprising two or more grasping methods where the outputs of two or more function approximators are combined to select between these grasping methods for a given sensor image.

60. The apparatus of claim 50, further comprising pushing objects to separate them from the environment and create an accessible grasp.

61. The computer-implemented method of claim 1, further comprising pushing objects to separate them from the environment and create an accessible grasp.

62. An apparatus, comprising:

a sensor;

a robotic targeting mechanism; and

one or more processors configured to use sensor images to compute a desired target configuration for the robotic targeting mechanism based at least in part on a function approximator that is trained on a training dataset of labelled examples of sensor images and target points using:

a set of three-dimensional (3D) models of objects;

one or more analytic mechanical representations of target points; and

statistical sampling to model uncertainty in either or both of sensing and control;

wherein the function approximator is configured to take as input a sensor image and return data that is used to compute robust target points for the robotic targeting mechanism.

63. The apparatus of claim 62, further comprising obtaining a 3D depth map using at least one selected from a group consisting of: a structured lighting system, a Lidar system, a stereo pair of color cameras, a stereo pair of monochrome cameras, and a monocular image.

64. The apparatus of claim 62, wherein the sensor image includes a 2D image from a camera.

65. The apparatus of claim 62, further comprising use of motion planning methods to avoid robot contact with the environment based on collision checking.

66. The apparatus of claim 62, further comprising means for detecting the outcome of target attempts.

67. The apparatus of claim 66, wherein detecting the outcome of target attempts includes using at least one selected from a group consisting of: one or more load cells, a light sensor, a camera, a force sensor, and a tactile sensor.

68. The apparatus of claim 62, further comprising applying multiple targeting methods in parallel.

69. The apparatus of claim 62, further comprising two or more targeting methods where the outputs of two or more function approximators are combined to select between these targeting methods for a given sensor image.

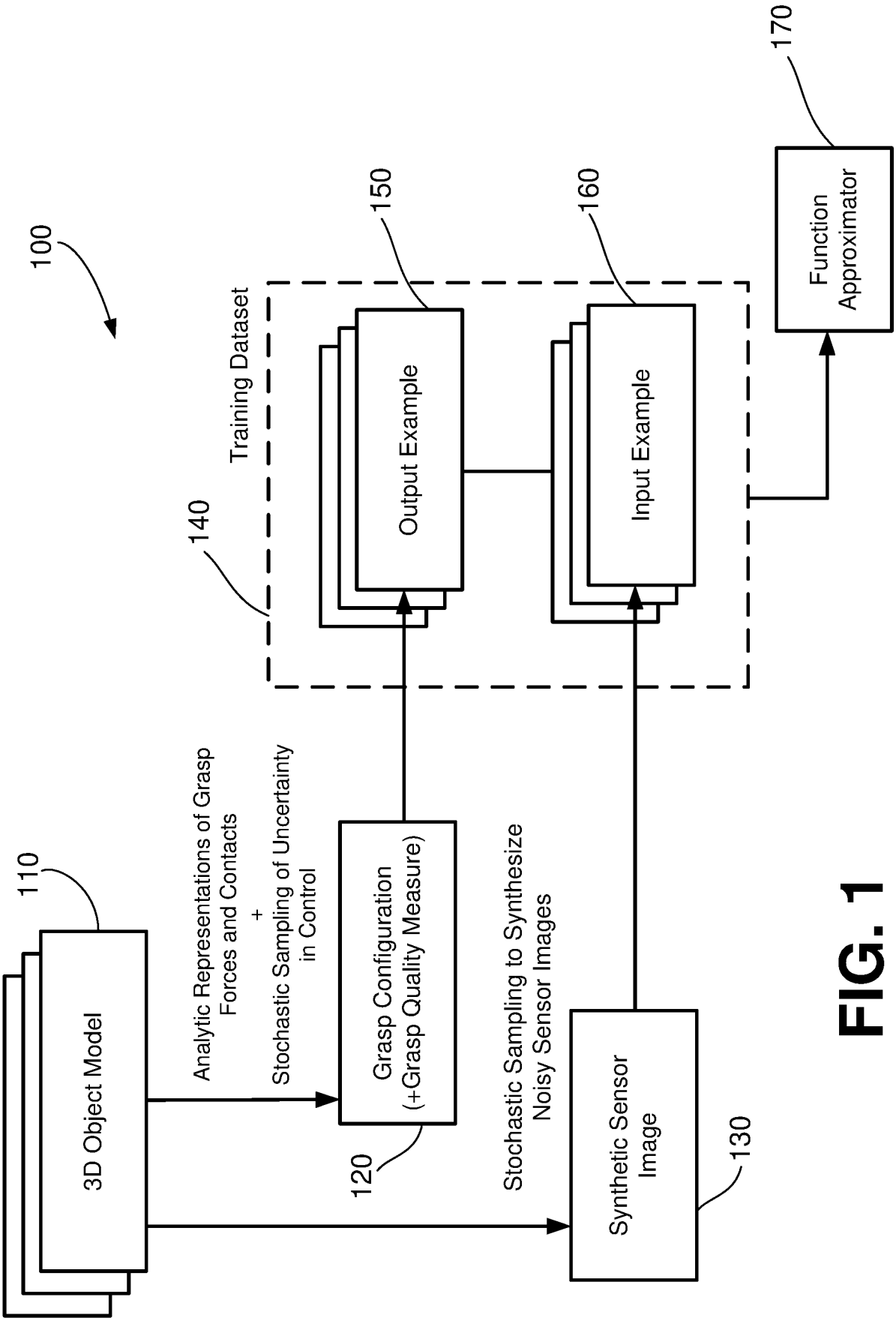
100
110
3D Object Model
Analytic Representations of Grasp
Forces and Contacts
+
Stochastic Sampling of Uncertainty
in Control
120
Grasp Configuration
(+Grasp Quality Measure)
Stochastic Sampling to Synthesize
Noisy Sensor Images
130
Synthetic Sensor
Image
140
Training Dataset
150
Output Example
160
Input Example
170
Function
Approximator

FIG. 1

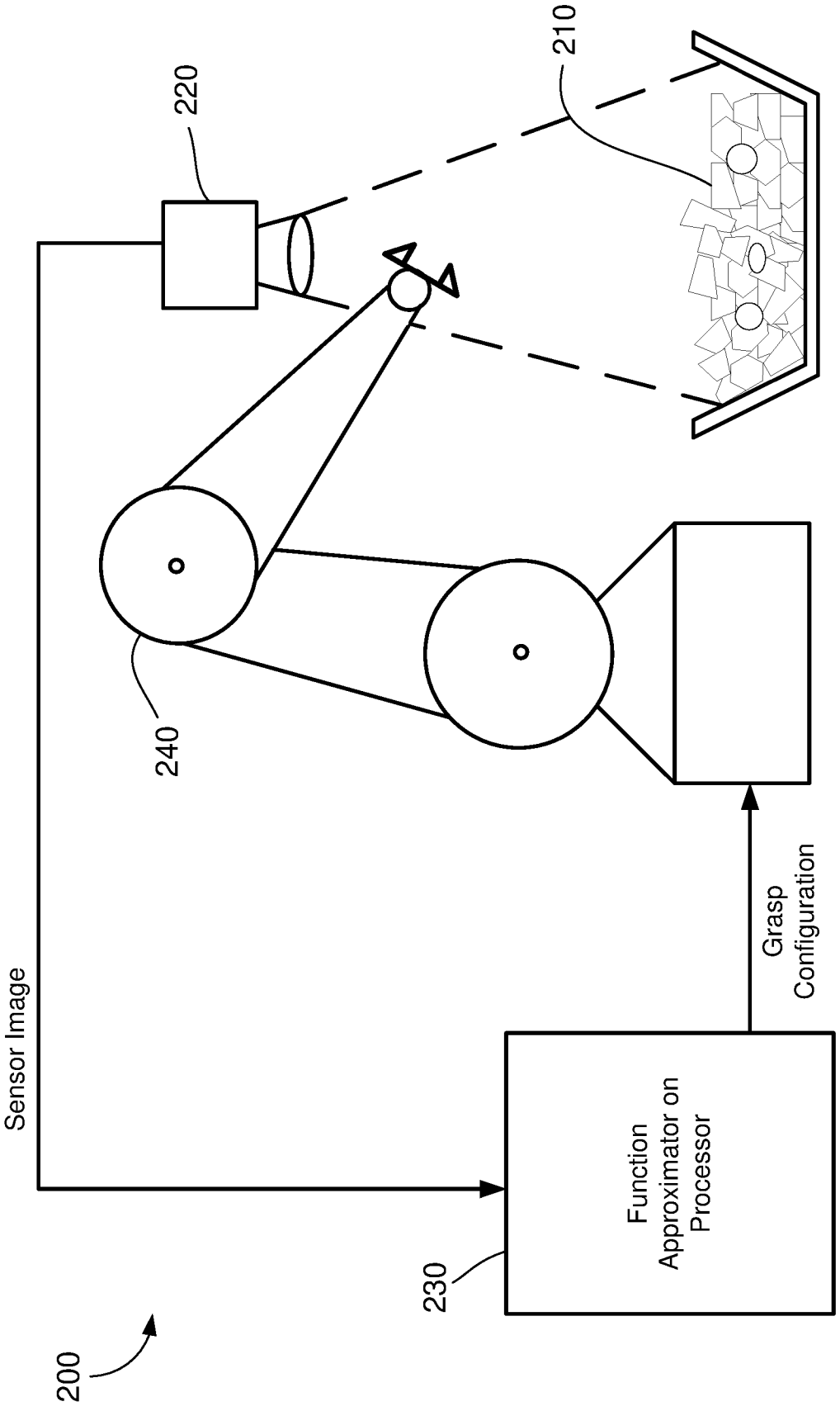
200
220
210
240
230
Function Approximator on Processor
Sensor Image
Grasp Configuration

FIG. 2