Data-Driven Grasp Synthesis - A Survey

Reminders

- dexterous = skillful (esp. with hands) – e.g. being able to move the pen around while it is grasped
- wrench = noncontact forces and moments applied to the object
- force closure $\implies$ grasp can be maintained in the face of any object wrench – is aided by friction (c.f. form closure means no movement at all)

Overview

- *grasp synthesis* = "problem of finding a grasp configuration that satisfies a set of criteria relevant for the grasping task"

1. Analytic
    - constrained optimization problem over some of the following criteria on force closure grasps
        - *dexterity*, *equilibrium*, *stability* or certain *dynamics*
    - not commonly used outside of simulation due to sensitivity to noise
2. Empirical (also called data-driven, comparative, knowledge-based)
    - sampling grasp candidates and ranking them according to a specific metric
    - usually parametrised by: the grasping point on the object, approach vector towards the grasping point, wrist orientation and initial finger configuration
    - cannot provide guarantees on analytic criteria, but are less robust to noise and generalize better

Key ingredients

- Problems with the analytic methods:

1. Relative position of object and the manipulator is only approximately known
    - somewhat solved using *independent contact regions* (regions of valid finger positions) and *caging formulations*
2. Precise geometric and physical models of objects are not available
    - particle filters for estimation, but still limited to simulation

- classic metrics (such as ϵ-metric) shown to not be good indicators of grasp success IRL
    - idea of learning grasps arised
- data-driven approach divided into grasps of known, familiar and unknown objects

Known objects

- Focus on object recognition and pose estimation

Offline generation of grasp experience database

1. 3D mesh models and contact-level grasping
    - assume 3D mesh is available $\rightarrow$ sample grasp hypotheses $\rightarrow$ rank
    - ranking with ϵ-metric shown bad results under uncertainty
2. Learning from humans
    - observe how humans grasp with cameras, magnetic tracking devices, special glowes etc.
    - imitate the results e.g. using reinforcement learning
3. Learning through trial and error
    - refine grasp candidates by trail and error, e.g. using reinforcement learning

Online object pose estimation

- Recognize the object and estimate its relative pose online → use the learned grasp database to perform the grasp of **the** object

Familiar objects

- Use similarity matching to a set of known objects
- Similarity can be texture, specific local features, functionality, etc. → difficult to find a representation for similarity

Discriminative approaches

- Learn to discriminate between good and bad grasps or graspable and not graspable parts of the object

1. Based on 3D data
    - use superquadrics (SQ) to approximate objects or parts of objects and learn if they are graspable
        - SQs nice as they have low number of parameters and high shape variability, but not sure how to deal with noisy data
    - use other approximations such as boxes, Markov Random Fields, etc.
2. Based on 2D data
    - avoid complexity of 3D data and mainly rely on 2D data to discriminate between good and bad grasp locations
    - use certain experience or imitate human interaction
    - works for certain applications, but in general highly under-constrained
3. Integrating 2D and 3D data
    - e.g. segmentation of local or global 2D shapes for learning potential grasp points

Grasp synthesis by comparison

- Grasp hypothesis synthesised from similar objects with known good grasps

1. Synthetic exemplars
    - low-level object features to encode similarity (e.g. shape, texture, weight), or semantic level features (e.g. object category, desired task)
    - some work addressing what to do in case 3D object meshes not available (e.g. estimating 3D mesh from sensor data)
2. Sensor-based exemplars
    - map real sensor data to grasps → potential for generalization by learning the mapping

Generative models for grasp synthesis

- These approaches identify common structures from examples, instead of discriminating grasps or comparing to previous examples
- Not much work done

Category-based grasp synthesis

- Previous approaches link low-level object information to grasp, with assumption that similar objects should be grasped the same way
- Objects might have a completely different shape and appearance, but be graspable the same way (e.g. based on functionality)

Unknown objects

- Focus on extracting features indicative of good grasps

Approximating unknown object shape

- Approximate objects with shape primitives (+ using symmetries and heuristics) and a grasp sampled

Grasp hypotheses from low-level features

- Map low-level 2D or 3D visual features to a set of grasp postures and rank them depending on the criteria
- E.g. segment object parts and grip the best shape for the gripper, or the shape with most interior points to the gripper

Grasp hypotheses from global shape

- Use certain heuristics based on full object shape to infer one good grasp hypothesis

Comments

- Four major areas that form open problems:

1. Object segmentation
    - General solution still an open problem. Different scenes, cluttered environments, etc.
2. Learning to grasp
    - Continuous learning about new objects, how to manipulate them, autonomously quantify good or bad grasp attempts, etc.
3. Autonomous manipulation planning
    - Problems more complex than grasping from a table top
4. **Robust execution**
    - Using constant visual or tactile feedback to adapt to unforseen situations
    - How can tactile feedback be used to correct the grasp irrespective of the object? How can tactile and visual information be fused?

- In my opinion a rather interesting survey outlining most important research in the field
- Cons: some parts left unclear, and some important terminology assumed to be known