Robotics and Autonomous Systems 62 (2014) 497–505

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Occlusion-aware multi-view reconstruction of articulated objects for manipulation

Xiaoxia Huang [a], Ian Walker [a], Stan Birchfield [a,b,*]

[a] Electrical and Computer Engineering Department, Clemson University, Clemson, SC 29634, United States
[b] Microsoft Research, Redmond, WA 98052, United States

HIGHLIGHTS

- Use of the Procrustes-Lo-RANSAC (PLR) algorithm to perform 3D reconstruction of articulated objects.
- A purely geometric approach recovers the axes' location, orientation, and type (revolute or prismatic).
- The resulting 3D model is occlusion-aware, meaning that parts not visible in the current view are included.

ARTICLE INFO

Article history:
Received 13 June 2012
Received in revised form
10 July 2013
Accepted 10 December 2013
Available online 22 December 2013

Keywords:
Articulated reconstruction
3D reconstruction
Procrustes analysis
Locally optimized RANSAC

ABSTRACT

We present an algorithm called Procrustes-Lo-RANSAC (PLR) to recover complete 3D models of articulated objects. Structure-from-motion techniques are used to capture 3D point cloud models of an object in two different configurations. Procrustes analysis, combined with a locally optimized RANSAC sampling strategy, facilitates a straightforward geometric approach to recovering the joint axes, as well as classifying them automatically as either revolute or prismatic. With the resulting articulated model, a robotic system is then able to manipulate the object along its joint axes at a specified grasp point in order to exercise its degrees of freedom. Because the models capture all sides of the object, they are *occlusion-aware*, meaning that the robot has knowledge of parts of the object that are not visible in the current view. Our algorithm does not require prior knowledge of the object, nor does it make any assumptions about the planarity of the object or scene. Experiments with a PUMA 500 robotic arm demonstrate the effectiveness of the approach on a variety of real-world objects containing both revolute and prismatic joints.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The transition of robots from constrained industrial environments to unstructured, dynamic environments is crucial for emerging application areas such as assistive and service robotics. For such applications, one important capability is that of reconstructing *articulated* objects [1–4], which are sets of rigid links connected by one or more (revolute or prismatic) joints. A surprisingly large number of important objects encountered every day can be modeled in this fashion, such as refrigerators, microwave ovens, drawers, doors, laptop computers, scissors, staplers, and so forth.

While considerable effort has been spent on recovering the joint parameters of a human subject from video or motion capture [5–9], reconstruction of non-human articulated objects has only recently begun to gain significant attention from researchers. Early work in this area focused on estimating joint axes by clustering tracked feature points, from which camera projection matrices were then recovered by assuming that the scene consists of planar surfaces rotating about vertical axes [10], or by imposing motion constraints on the projection equations [11].

In the computer vision community, factorization methods for affine reconstruction [12,13] are perhaps the most popular approach to articulated reconstruction. For example, by adding articulation constraints to the traditional formulation, Tresadern and Reid [14] detect the articulated objects, determine their degrees of freedom, and locate the joints. Similarly, Paladini et al. [15] recover 3D shape and motion of non-rigid and articulated objects in the case of missing data using an iterative factorization approach. Yan and Pollefeys [16] also investigate the subspace properties of articulated motion in a factorization framework by segmenting feature trajectories by local sampling and spectral clustering, then building the kinematic chain as a minimum spanning tree of a graph constructed from the segmented motion subspaces. More recent work by Fayad et al. [17] uses a hill-climbing approach that minimizes a single energy functional based

* Corresponding author at: Electrical and Computer Engineering Department, Clemson University, Clemson, SC 29634, United States.
E-mail addresses: xhuang@clemson.edu (X. Huang), iwalker@clemson.edu (I. Walker), stb@clemson.edu (S. Birchfield).

0921-8890/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.robot.2013.12.006

on image reprojection error, with alternating steps utilizing graph cuts to assign points to links, then factorization to reconstruct 3D models of the links. Note that unlike these factorization-based approaches, which are limited to affine projection, our system allows for perspective projection.

In robotics, several methods have been proposed to reconstruct articulated objects with unknown skeletal parameters. Sturm et al. [4] recover kinematic models of 1-DOF articulated objects such as a microwave by tracking the poses and orientations of rigid parts captured by a motion capture system and addressing a mixture of parameterized and parameter-free (Gaussian process) representations to best explain the given observation. In follow up work, the same researchers [3] learn articulation models of objects without using artificial markers by applying generative models to depth images obtained from a self-developed active stereo system. In contrast to their work, our approach is not restricted to planar objects. Similar work by Katz et al. [1] reconstructs 3D kinematic structures of rigid articulated bodies in a single-view using feature tracking, motion segmentation, and structure-from-motion techniques.

One limitation of current approaches to articulated object reconstruction is their restriction to processing data from a single view. As a result, such approaches do not yield any information about the occluded portion of the object (that is, the back side that is not visible in the current view), thus preventing manipulation of these non-visible portions. We introduce the term *occlusion-aware* to refer to a model's ability to capture knowledge of parts of the object that are not visible in the current view. We argue that in a robotics context it is important to be able to reason about non-visible portions of the object in order to manipulate them. For example, a robot might approach a stapler from a different direction during manipulation than it did during model construction. Such full 3D knowledge has always been assumed in the context of grasping research based on 3D CAD models [18–20].

Motivated by recent developments in automated reconstruction of complete 3D models from a collection of images [21–24], in this paper we present an occlusion-aware system for reconstructing articulated objects from images taken by a camera from different viewpoints. Our method, called Procrustes-Lo-RANSAC, or PLR, first builds two complete 3D point cloud models by applying structure-from-motion algorithms to images captured of the object in two different configurations. Then the method uses Procrustes analysis combined with a locally optimized RANSAC sampling strategy to automatically segment the points into the individual links. After aligning the links, the articulated structure of the object is estimated using a geometric approach. With hand–eye calibration, the robot can then align its coordinate system with that of the recovered articulated model and manipulate the object by exercising the degrees of freedom captured by the model. Unlike some previous techniques, our approach uses perspective (rather than affine) projection, and it does not make any planar assumptions about the scene. This paper is based on our earlier work in [25], extending it to the manipulation of articulated objects based on their models. We show the results of the system on a variety of everyday objects, demonstrating the effectiveness of the approach.

2. Learning articulated objects

Assume we have an object composed of a set of rigid links connected by revolute or prismatic joints, so that a *configuration* of the object refers to a specific set of values for the joint angles or displacements, as appropriate. Let $\bar{\mathbf{p}} = (\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(m)})$ and $\bar{\mathbf{q}} = (\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(m)})$ be two point clouds of the object in two different configurations, where $\mathbf{p}^{(i)} \leftrightarrow \mathbf{q}^{(i)}$, $i = 1, \ldots, m$, are corresponding points, and $\mathbf{p}^{(i)}, \mathbf{q}^{(i)} \in \mathbb{R}^3$. The heart of our PLR algorithm is to

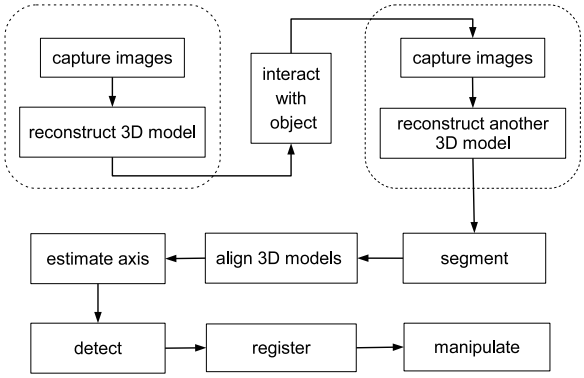

Fig. 1. Overview of our system.

assign a label $\lambda^{(i)} \in \mathbb{Z}^+$, $i = 1, \ldots, m$, to each point indicating to which link it belongs, as well as to compute the joint parameters $\{\mathbf{u}_{ab}, \omega_{ab}\}$ for each adjacent pair of links a and b, where $\mathbf{u}_{ab} \in \mathbb{R}^3$ is the free vector specifying the direction of the joint axis, and $\omega_{ab} \in \mathbb{R}^3$ is a point on the joint axis (specifically the average projection of all points on both links onto the axis). The point clouds define the 3D model of the object obtained by structure-from-motion, and correspondence is established between them using descriptors associated with the points. An overview of our system is illustrated in Fig. 1.

We assume that the capability of performing sufficient exploratory interaction with the object to change its configuration is present. In this way, the approach bears some resemblance to interactive perception [26,1,27–29], except that we allow either a human or robot to perform the interaction. Automatically planning the end effector motion path for interactive perception in such situations remains an unsolved problem, because some sort of preliminary model (at least) is needed in order to interact with the object, but the interaction is necessary to estimate the model. Therefore, having the user perform the interaction enables us to escape this difficult chicken-and-egg problem. As progress is made toward developing such autonomous exploratory behavior, the reconstruction method described in this paper still applies.

2.1. Building initial 3D models

The first step of the approach is to build 3D models of the object in two different configurations. (Note that only two configurations are needed, no matter how many links or joints.) With the object in each configuration, feature points are detected and matched in a set of images taken from different viewpoints. We use the Bundler algorithm [30,21], which calibrates the camera and computes the camera locations in 3D. With this information, patch-based multi-view stereo (PMVS) [31,22] is used to reconstruct dense 3D oriented points, where each point has an associated 3D location, surface normal, and a set of visible images. (See Fig. 2.)

2.2. Rigid link segmentation

Once the models have been constructed, the oriented 3D points of the models are segmented into the constituent rigid components of the object. We use the affine SIFT (ASIFT) feature detector [32], which is an affine invariant extension of the popular SIFT feature detector [33], to find features in every image of the two sets. For every feature point in an image of the first configuration, the matching feature point in the second configuration is found as the one that minimizes the sum-of-squared differences (SSD) between gray-level patches surrounding the two features. The same matching algorithm is then run in the reverse order by swapping the roles of the images, and matches are retained if

Download English Version:

https://daneshyari.com/en/article/412340

Download Persian Version:

https://daneshyari.com/article/412340

Daneshyari.com