Journal of Visual Languages and Computing 27 (2015) 38–48

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

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

Intermodal image-based recognition of planar kinematic mechanisms ☆

Matthew Eicholtz, Levent Burak Kara *

Department of Mechanical Engineering Carnegie Mellon University, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 3 October 2014
Received in revised form
20 October 2014
Accepted 21 October 2014
Available online 31 October 2014

Keywords:
Computer vision
Evolutionary multiobjective optimization
Kinematic simulation
Object recognition

ABSTRACT

We present a data-driven exploratory study to investigate whether trained object detectors generalize well to test images from a different modality. We focus on the domain of planar kinematic mechanisms, which can be viewed as a set of rigid bodies connected by joints, and use textbook graphics and images of hand-drawn sketches as input modalities. The goal of our algorithm is to automatically recognize the underlying mechanical structure shown in an input image by leveraging well-known computer vision methods for object recognition with the optimizing power of multiobjective evolutionary algorithms. Taking a raw image as input, we detect pin joints using local feature descriptors in a support vector machine framework. Improving upon previous work, detection confidence depends on multiple context-based classifiers of varying image patch size and greedy foreground extraction. The likelihood of rigid body connections is approximated using normalized geodesic time, and NSGA-II is used to evolve optimal mechanisms using this data. The present work is motivated by the observation that textbook diagrams and hand-drawn sketches of mechanisms exhibit similar object structure, yet have different visual characteristics. We apply our method using various combinations of images for training and testing, and the results demonstrate a trade-off between solvability and accuracy.

1. Introduction

The design of complex mechanical linkages is a challenging task involving the coordination of multiple rigid bodies to achieve a desired dynamic profile (see Fig. 1 for examples). The ability to visualize the kinematics of a mechanism is a valuable skill to improve mechanical intuition during design analysis and synthesis [1], yet current simulation tools may be insufficient for fast kinematic visualization. Currently, engineers will likely resort to one of three options. First, they may use mental simulations to infer mechanical behavior [2], but this is ineffective for people with low spatial ability [3] and is generally difficult for complex mechanisms [4]. Second, specialized software [5–6] may be used for simulations, but this task is often too time-consuming to be practical (e.g. students solving a dynamics homework problem, professional engineers brainstorming potential design concepts) and may require advanced programming skills, which hinders novice users. Third, engineers often use hand-drawn sketches to convey design ideas and visualize dynamic properties, perhaps abstracting the mechanism to a simpler form or using key annotations and arrows to demonstrate motion.

In a previous work [7], we developed an algorithm to bridge the gap between ineffective mental simulations and impractical computer simulations by automatically recognizing the underlying mechanical structure in

☆ This paper has been recommended for acceptance by S.-K. Chang.

* Corresponding author. Tel.: +1 412 268 2509.
E-mail addresses: meicholt@andrew.cmu.edu (M. Eicholtz), lkara@cmu.edu (L. Burak Kara).

http://dx.doi.org/10.1016/j.jvlc.2014.10.024
1045-926X/

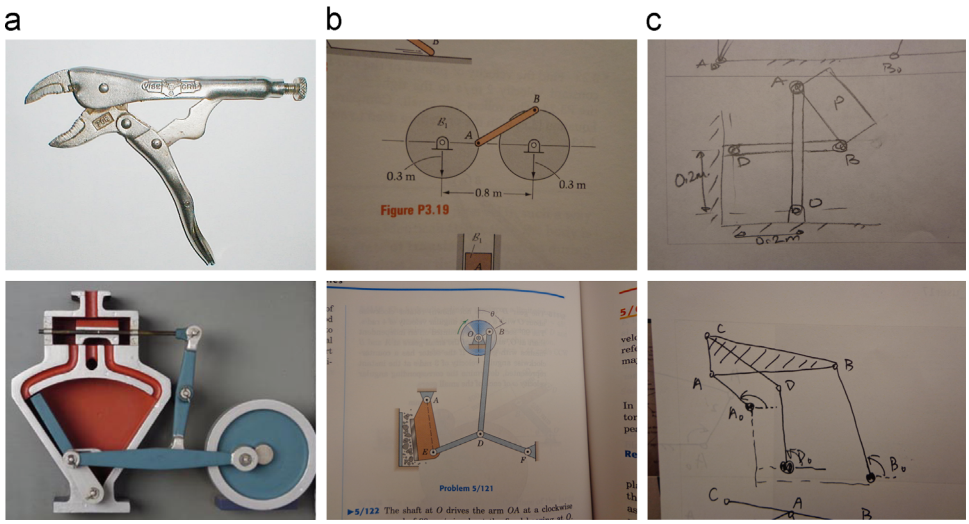

Fig. 1. Example mechanisms in (a) natural images of real-world objcets, (b) textbook graphics, and (c) hand-drawn sketches. Each mechanism contains a set of rigid bodies connected by kinematic pairs (e.g. revolute or prismatic joints) that constrain their motion. The present work focuses on automatically recognizing the number, location, and connectivity of joints in textbook graphics and hand-drawn sketches; this information is all that is required to fully specify the allowable motion of each rigid body.

a single image. At the heart of our approach was a novel combination of vision-based object recognition with multiobjective evolutionary opti-mization. The fundamental principle of the method was to consider mechanisms as a collection of connected joints, where each pairwise joint connection indicated that two joints were fixed to the same rigid body. We limited our study to planar mechanisms, in which the motion of every rigid body is constrained to the plane perpendicular to the viewer, and only considered examples made up entirely of revolute joints. With this representation, the task involved locating probable joints in an image using a sliding window object detector, assessing the likelihood of all pairwise joint connections using normalized geodesic time and maximizing image consistency and mechanical feasibility using the NSGA-II algorithm. The algorithm enabled the evolution of a small set of feasible mechanical structures based on local features in a single image, and only required a set of training images for joint detection. We initially implemented the approach on textbook graphics due to their relative simplicity and wide availability.

In the present work, outlined in Fig. 2, we shift our focus to include sketches as valid input data to our algorithm. This is motivated by the idea that sketches are more directly related to design synthesis than textbook graphics. Someone creating a new mechanism may not be able to find a clean image depicting their design concept; indeed, they may not even know what they are looking for yet. With our technology, we hope to enable users to rapidly explore the design space using pencil and paper without being encumbered by existing designs.

We represent sketch data as an image, so that no modifications to the original algorithm are explicitly required to accommodate the new input modality. Regardless, we propose a couple key enhancements to the joint detection scheme in order to boost performance; details are provided later in this paper. Despite being of the same "form" as the textbook images used previously, we still consider sketches to be from a different *modality* because they were created in a different manner than textbook graphics. The evidence in support of this proposition is clear from the examples pictured in Fig. 1. Textbook graphics use consistent shapes, colors, and textures, while sketches are typically messier, have curvier lines, and include artifacts such as overtracing, tonal variation in stroke intensities, and cross hatching, among others [8]. Furthermore, depictions of mechanisms in textbooks may be surrounded by irrelevant text, annotations, highlighting, or other mechanisms that clutter the image; sketches, on the other hand, can be created without such distracting visual elements.

Even though they may be strikingly different in certain visual characteristics, textbook graphics and sketches of mechanisms adhere to the same structural principles. This poses an interesting problem: can we successfully use one input modality for training and the other for testing? More specifically, are we required to have a set of training sketches in order to correctly recognize test sketches of mechanisms? The answer may have important implications for future tools involving the recognition of visual objects with different input modalities.

The remainder of the paper is structured as follows: section 2 highlights related work in sketch recognition, computer vision, and evolutionary algorithms. Improvements made to our original algorithm are provided in section 3. Experimental methods, including results and discussions, are given in section 4, followed by concluding remarks in section 5.

2. Related work

2.1. Object detection

Object detection is a mature field of research in computer vision, spanning countless real-world applications. A typical object detector extracts salient features