Computers & Graphics ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

Characterizing the performance of an image-based recognizer for planar mechanical linkages in textbook graphics and hand-drawn sketches ☆

Matthew Eicholtz, Levent Burak Kara *

Department of Mechanical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 9 March 2015
Received in revised form
1 June 2015
Accepted 7 June 2015

Keywords:
Computer vision
Evolutionary multiobjective optimization
Image processing
Kinematic modeling
Object recognition
Sketch recognition

ABSTRACT

In this work, we present a computational framework for automatically generating kinematic models of planar mechanical linkages from raw images. The hallmark of our approach is a novel combination of supervised learning methods for detecting mechanical parts (e.g. joints, rigid bodies) with the optimizing power of a multiobjective evolutionary algorithm, which concurrently maximizes image consistency and mechanical feasibility. A rigorous set of experiments was conducted to systematically evaluate the performance of each phase in our framework, comparing various combinations of joint and body detection schemes and feasibility constraints. Precision–recall curves are used to assess object detection performance. For the optimization, in addition to standard accuracy measures such as top-*N* accuracy, we introduce a new performance metric called *user effort ratio* that quantifies the amount of user interaction required to correct an inaccurate optimization solution. Current state-of-the-art performance is achieved with (i) one (or a cascade of) support vector machines for joint detection, (ii) foreground extraction to reduce false positives, (iii) supervised body detection using normalized geodesic time, distance, and detected joint confidence, and (iv) feasibility constraints derived from graph theory. The proposed framework generalizes moderately well from textbook graphics to hand-drawn sketches, and *user effort ratio* results demonstrate the potential power of an interactive system in which simple user interactions complement computer recognition for fast kinematic modeling.

1. Introduction

A planar mechanical linkage is an assembly of rigid bodies connected by kinematic pairs (or joints) that constrain its motion within a plane. With applications in robotics [6,7], healthcare [29,38], transportation [58,65], and industrial equipment [32,46], among others, we observe and make use of the dynamic behavior of complex mechanical systems on a daily basis. Visualizing the coordination motion of mechanical linkages is indeed a valuable skill for improving design intuition [18], yet during the design and analysis of such dynamic assemblies, the visual content is largely static in nature, as illustrated in Fig. 1.

To overcome this information deficit, students and engineers may use mental simulations to infer mechanical behavior [36], but this can be difficult for complex problems [35] or for individuals with low spatial ability [37]. They also frequently use hand-drawn sketches to convey design ideas, perhaps incorporating key annotations and arrows to demonstrate motion; even so, the burden for visualization remains with the user. Alternatively, computer simulations can be generated using specialized software [1–4], in which users manually create kinematic models. However, 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, hindering novice users. There is a clear need for better software tools that facilitate quick computer-based kinematic visualization of mechanical linkages, filling the gap between ineffective mental simulations and impractical manual model creation.

In this paper, we address that need by proposing a computational framework that automatically recognizes the underlying mechanical structure in images of textbook graphics or hand-drawn sketches. In order to generate a proper kinematic simulation of a planar mechanical linkage, the user typically needs to specify the number and position of rigid bodies that make up the linkage, as well as the type and location of joints that dictate relative motion between rigid bodies. The goal of our method is to offload this burden to the computer. We accomplish this in

☆This article was recommended for publication by F. Samavati.

* Corresponding author.
E-mail addresses: meicholt@andrew.cmu.edu (M. Eicholtz), lkara@cmu.edu (L.B. Kara).

http://dx.doi.org/10.1016/j.cag.2015.06.002
0097-8493/

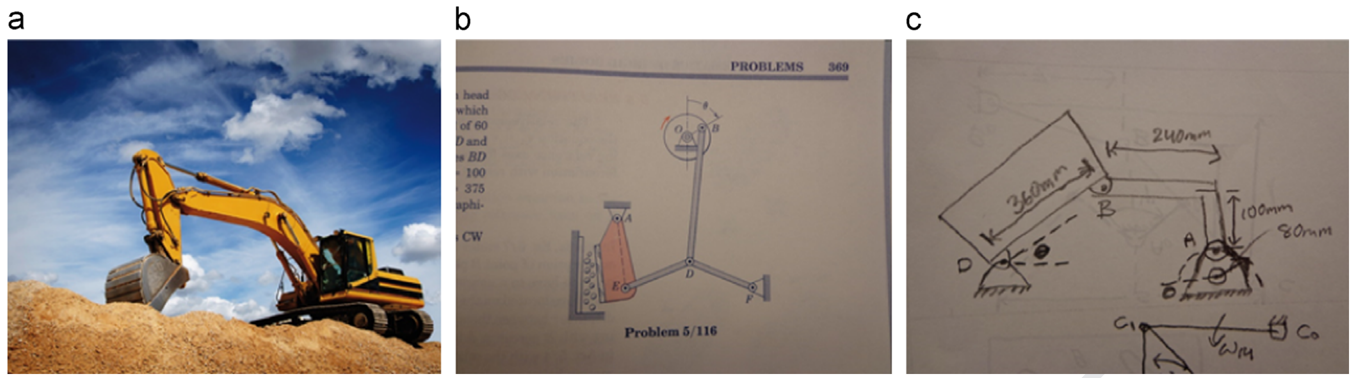

Fig. 1. Example mechanical linkages (a) in the real world, (b) in textbooks, and (c) in hand-drawn sketches. Ironically, even the excavator shown here is static in this printed document, despite our usual observation of its dynamic behavior in the real world. Images courtesy of (a) MOBA (www.moba.de), (b) the MECH135 dataset, and (c) the MECHS250 dataset. The latter two cases represent valid inputs for the recognition framework presented in this paper.

multiple stages using a joint-centric approach, meaning that linkages are viewed as a collection of connected joints (as opposed to connected bodies), and each pairwise joint connection indicates that those two joints exist on the same rigid body. In this way, rigid bodies can be inferred from the connections between joints. This reduces the problem to localizing joints in the image, predicting which joints coexist on rigid bodies, and resolving discrepancies to form reasonable mechanical assemblies. For simplicity, we limit our study to planar mechanisms comprising only revolute (pin) joints.

2. Background

Our computational approach relies on ideas from many disciplines, including computer vision, sketch recognition, and evolutionary computation. Here, we outline key references in these areas that influenced the design of our recognition framework.

2.1. Scene understanding

The task of identifying structured objects in images is not a new one [14,24,25]. Practical applications include face recognition [26,66], pose estimation [64], and 3D surface estimation [55]. The key difference, though, between previous work in this area and our present domain is that planar mechanisms do not have well-defined structural or spatial dependencies. For example, in face recognition, it is straightforward to learn that a chin should not be located above the nose or that eyes should exist between the ears; with mechanical linkages, it is less clear if a specific joint should be systematically connected to another. Little knowledge is gained about the likelihood of other objects in the image just from knowing one object's location.

Within the current domain of interest, Sato et al. [57] proposed a vision-based approach for automatically estimating the location of an axis of rotation in a mechanical linkage. The primary differences between that work and the research in this paper are twofold. First and foremost, it relies on motion tracking from a series of images to capture the moving parts, whereas our work is restricted to a single image. Second, it seems to be limited to simple mechanisms with only one axis of rotation (single pin joint), whereas our approach can handle more complex kinematic behavior (multiple pin joints).

2.2. Sketch recognition

Two important aspects of sketch recognition that relate to the present work are representation and complexity. With regard to representation, two classes of techniques have emerged in the literature. Stroke-based methods treat each sketch as a sequence of time-stamped strokes, each containing a series of sample points in space. While some works share similarities to our domain [17,33,34,42,52], stroke-based methods are ill-suited for our recognition framework, which is designed to work on rasterized images. Still, there are interesting parallels; for instance, [34] uses a graph representation to combine low-level primitives into high-level shapes using geometrical rules. We also implement graphs in our recognition pipeline, but instead connect low-level joints to form high-level mechanisms based (partially) on mechanical feasibility rules.

The other class of sketch recognition techniques is image-based approaches, including the present work, which neglect temporal information and only consider the spatial layout of pixels. This poses the additional challenge of grouping relevant pixels, depending on the object being recognized. With regard to sketch complexity, it is important to distinguish between isolated symbol recognizers and detecting objects in freehand sketches, which is a more challenging problem. The task of symbol recognition can be treated as a template matching problem; some examples of successful approaches in this area include [13,28,39,40,43,50]. In some sense, the joint recognition algorithm used here is similar to a sliding window symbol recognizer. However, due to the allowable shape variance of objects in mechanical linkages, we do not use unsupervised part templates and instead learn a discriminative model based on local image features.

Within the current domain of interest, researchers have briefly studied the automatic recognition of mechanical systems from sketch input [17,27,28], but these approaches typically involve clean images, well-defined part templates, and sometimes make use of temporal information to aid recognition. Our approach must generalize well to raw images, which are often noisy, may contain extraneous information from other graphics, and do not always contain well-defined part models.

2.3. Evolutionary multiobjective optimization

The proposed framework includes an evolutionary optimization stage to resolve discrepancies from the vision-based detection of joints and joint connections. There is a growing body of research in the area of multiobjective evolutionary algorithms (MOEAs), especially in regard to real-world applications. Many well-known MOEAs are based on Pareto dominance [19,61,67], which states that a given solution dominates another solution if it is at least as good on all objectives and better on at least one objective. Arguably the most popular MOEA of this type and the one used in this present work is the nondominated sorting genetic

Please cite this article as: Eicholtz M, Kara LB. Characterizing the performance of an image-based recognizer for planar mechanical linkages in textbook graphics and hand-drawn sketches. Comput Graph (2015), http://dx.doi.org/10.1016/j.cag.2015.06.002