Computers & Graphics 36 (2012) 945–954

Contents lists available at SciVerse ScienceDirect

Computers & Graphics

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

Technical Section

Real-time interactive modeling and scalable multiple object tracking for AR ☆

Kiyoung Kim [a], Vincent Lepetit [b], Woontack Woo [c,*]

[a] GIST U-VR Lab., South Korea
[b] EPFL CV Lab., Switzerland
[c] KAIST GSCT UVR Lab., South Korea

ARTICLE INFO

Article history:
Received 25 April 2012
Received in revised form
7 July 2012
Accepted 16 July 2012
Available online 23 August 2012

Keywords:
Augmented Reality
Structure from Motion
Multiple object tracking
Interactive modeling

ABSTRACT

We propose a real-time solution for modeling and tracking multiple 3D objects in unknown environments for Augmented Reality. The proposed solution consists of both scalable tracking and interactive modeling. Our contribution is twofold: First, we show how to scale with the number of objects using keyframes. This is done by combining recent techniques for image retrieval and online Structure from Motion, which can be run in parallel. As a result, tracking 50 objects in 3D can be done within 6–35 ms per frame, even under difficult conditions for tracking. Second, we propose a method to let the user add new objects very quickly. The user simply has to select in an image a 2D region lying on the object. A 3D primitive is then fitted to the features within this region, and adjusted to create the object 3D model. We demonstrate the modeling of polygonal and circular-based objects. In practice, this procedure takes less than a minute.

1. Introduction

Considerable progress has been made recently for camera tracking in unknown environments, in terms of reliability [1] and scalability with the size of the scene [2]. By contrast, tracking of 3D objects moving independently received much less attention, despite its importance for Human Computer Interfaces or Augmented Reality (AR). It often based on an offline stage during which the object 3D models and some additional information are collected, but this takes much time and requires additional knowledge.

We therefore propose a real-time solution of modeling unknown objects and tracking them independently as shown in Fig. 1. Our contribution is twofold. First, we show how to handle a large number of objects. For example, we can maintain real-time performances while simultaneously tracking 14 different objects, while continuously checking for objects from a database of 50 objects. To do that, we built upon the approach developed in [3], which is able to manage a large database of planar objects but track only one object at a time. Object detection and feature tracking run in parallel: a foreground thread tracks feature points from frame-to-frame to ensure real-time performances, while a background thread aims at recognizing the visible targets and estimating their poses. This is done by efficiently matching the input frames against a database of registered keyframes of the target objects. We show how to detect all the known objects present in an image to be able to track reliably several objects simultaneously.

The second part of our contribution is to show how to allow the user to add new objects very efficiently with recovery of their shapes and the required feature data. In a way similar to [1], a second background thread maps the environment to track the camera and reconstruct the 3D locations of image features. When the user wants to define a new object, he/she simply has to select from a captured frame a region that lies on the object. We then fit automatically a 3D primitive to the 3D locations corresponding to the image features that belong to this region, to obtain the object 3D model. This procedure is fast and intuitive, and allows the creation of reliable 3D models and data required for tracking with a minimal effort.

This paper overcomes the limitations of the previous work [4] in the detection and the modeling. We extend the multiple object detection to be robust against erroneous keyframe searching results. The proposed method uses multiple keyframes instead of a single one per object to obtain many inliers. It slightly increases the detection time while it provides higher chance to get reliable poses in the tracking thread. We also enhance the modeling interface to support polygonal or circular-based models with a fast keyframe generation while the previous work only supported box-type objects. Additionally the post-processing using edges during the modeling process helps users to refine the initial inputs.

In the remainder of the paper, we first address the related work of the multiple object tracking and the interactive modeling in Section 2. We then explain our approach: real-time tracking

☆This article was recommended for publication by Associate Editor Mark Billinghurst.

* Corresponding author at: KAIST GSCT UVR Lab., South Korea.
Tel.: +82 42 350 2923; fax: +82 42 350 2910.
E-mail address: wwoo@kaist.ac.kr (W. Woo).

0097-8493/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2012.07.003

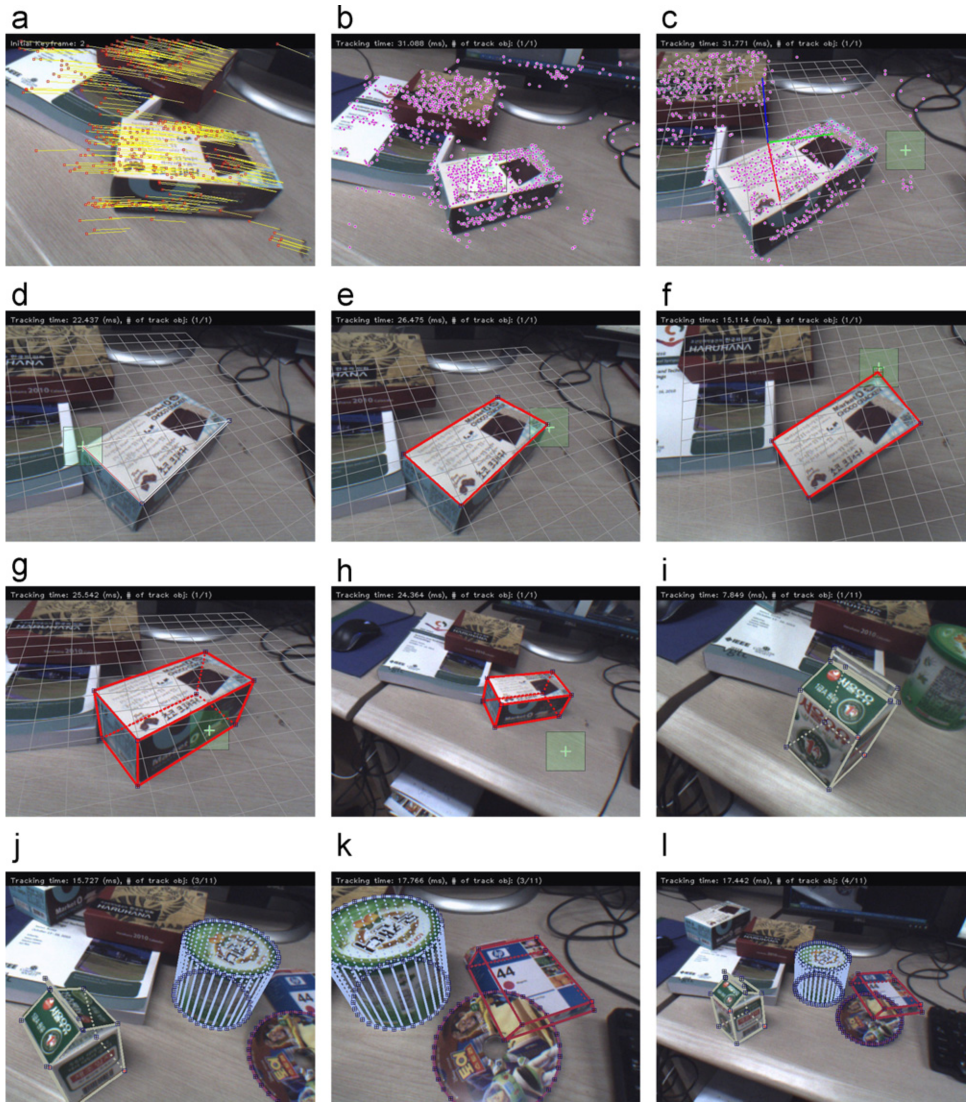

Fig. 1. Overview of the proposed approach. (a) Our system continuously reconstructs the environment and tracks the camera from the video stream. (b and c) The user can set a local coordinates upon the object surface by clicking once. (d and e) Then, the user can add control points at each corner of the object to define one facet. (f) The control points can be adjusted from another point of view. (g) If the object geometry is a prism, its 3D model can be quickly obtained by extending the facet to a 3D volume. Then, (h) the object can be tracked independently. (i–l) Multiple object tracking. Various types of 3D objects are modeled and tracked very fast.

and modeling in Section 3. The experimental results are presented in Section 4. Finally, we conclude this paper in Section 5.

2. Related work

Below, we review previous work on multiple object tracking and interactive modeling. Note that "scalable" term in this paper is about the number of target objects that can be managed in the tracker.

2.1. 3D Multiple object tracking

Park et al. [5,6] proposed a real-time tracking algorithm for multiple 3D objects that combines detection and frame-to-frame tracking. The main drawback of the approach is that its complexity in terms of computation time and memory grows linearly with the number of objects in a database, which makes it impractical when considering more than 10 objects. Wagner et al. [7] optimizes a score evaluating the trade-off between detection and frame-to-frame tracking, as detection is more robust but requires more time, to dynamically adjust the tracking and detection loads at run-time. It also "masks" the parts of the image where an object is already tracked, to avoid running the detection process on these parts and thus save computation time. It can handle a remarkably large number of objects simultaneously visible, however it is still limited in the number of objects that can potentially become visible, in other words the number of known objects in the database.

Very recently, Pilet and Saito [8] developed an approach that can scale to a very large number of known planar objects, by relying on image retrieval techniques similar to the one we use [9]. In this paper, we show how to combine detection and tracking for multiple non-planar objects, so that these two parts of the approach can be run in parallel and communicate when needed. Thus, our method runs at more than 30–60 Hz on a modern PC, while [8] reports about 8 Hz.

We also propose a method to let the user add new 3D objects very quickly, as discussed in the next subsection.

2.2. Online interactive modeling

Early works on online interactive modeling include [10,11]. Freeman and Steed [10] rely on a visual marker while we do not