Computers in Industry 99 (2018) 53–68

Contents lists available at ScienceDirect

Computers in Industry

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

Portable 3-D modeling using visual pose tracking

Klaus H. Strobl*, Elmar Mair[1], Tim Bodenmüller, Simon Kielhöfer[2], Tilo Wüsthoff, Michael Suppa[3]

Robotics and Mechatronics Center, German Aerospace Center (DLR), D-82230 Wessling, Germany

ARTICLE INFO

Article history:
Received 17 August 2017
Received in revised form 8 March 2018
Accepted 15 March 2018
Available online xxx

Keywords:
3-D modeling
Pose tracking
SLAM
Visual odometry

ABSTRACT

This work deals with the passive tracking of the pose of a close-range 3-D modeling device using its own high-rate images in realtime, concurrently with customary 3-D modeling of the scene. This novel development makes it possible to abandon using inconvenient, expensive external trackers, achieving a portable and inexpensive solution. The approach comprises efficient tracking of natural features following the Active Matching paradigm, a frugal use of interleaved feature-based stereo triangulation, visual odometry using the robustified V-GPS algorithm, graph optimization by local bundle adjustment, appearance-based relocalization using a bank of parallel three-point-perspective pose solvers on SURF features, and online reconstruction of the scene in the form of textured triangle meshes to provide visual feedback to the user. Ideally, objects are completely digitized by browsing around the scene; in the event of closing the motion loop, a hybrid graph optimization takes place, which delivers highly accurate motion history to refine the whole 3-D model within a second. The method has been implemented on the DLR 3D-Modeler; demonstrations and abundant video material validate the approach. These types of low-cost systems have the potential to enhance traditional 3-D modeling and conquer new markets owing to their mobility, passivity, and accuracy.

1. Introduction

3-D modeling already assumed a central role in areas like industrial inspection and recognition, reverse engineering, cultural heritage, medical imaging, computer graphics, and robotics. Other areas like leisure gaming, human–computer interaction, robotics, forensics, agriculture, and construction show a less direct requirement for 3-D modeling, but are increasingly taking advantage of it as a means to solve the visual perception problem. Visual perception is the process by which visual sensory information about the environment is received and interpreted; it is believed that it is through the explicit formation of 3-D models that a considerable number of the challenges on visual perception will be eventually solved. This is, of course, subject to the performance, flexibility, and cost of 3-D modeling devices.

Several factors like object self-occlusion, object size, or limited field of view make it impossible for a 3-D modeling system to acquire a complete model in a single measurement step, especially in close-range. Multiple views (or multiple sensors) are required to merge data to a single model. The prevalent approach is to measure the position and orientation (pose) of the sensor while acquiring range data, thereby registering multiple views into the same frame of reference. A range of tracking systems, robotic manipulators, passive arms, turntables, CMMs, or electromagnetic devices are deployed for this purpose. These options are inconvenient for three reasons: *First*, they limit mobility; *second*, they require accurate synchronization and extrinsic calibration (and cannot be rearranged); *third*, they usually represent the largest and most expensive part of the 3-D modeling system.

In this work we present an overview on the state of the art of close-range 3-D modeling systems regarding their data registration concept. We then make the case for data registration by visual pose tracking in realtime and go on describing their adaption for close-range 3-D modeling devices, using the video captured by their own cameras. Cameras are preferred sensors in many areas because they are light, affordable, consume less energy, allow for a very accurate parametrization of its operating model, and still they gather a plethora of information (both radiometric and geometric) within a single, rapid measurement. Further benefits exist:

---

* Corresponding author.
E-mail address: klaus.strobl@dlr.de (K.H. Strobl).
URL: http://rmc.dlr.de (K.H. Strobl)

[1] Present address: X Development LLC, 1600 Ampitheatre Pkwy, Mountain View, CA 94043, U.S.

[2] Present address: Ringstraße 38, D-86911 Dießen am Ammersee, Germany.

[3] Present address: Roboception GmbH, Kaflerstr. 2, D-81241 Munich, Germany.

https://doi.org/10.1016/j.compind.2018.03.009

cameras are non-contact sensors, thus free-floating, and passive since they do not need to project or exert action on the environment. In addition, visual pose tracking becomes inherently calibrated and synchronized with further image-based sensing.

Visual pose tracking is a hard problem because geometric information becomes entangled in radiometric and perspective geometric issues. Following distinct regions of interest in the images in realtime (**feature-based tracking**) is a popular technique to overcome this problem. This is especially demanding in close-range because features move faster than in medium- or long-range because they are also affected by camera translation. We proposed two novel schemes for efficient feature tracking on this type of devices: either leveraging an inertial measurement unit (IMU) [1] or adopting the Active Matching paradigm in Ref. [2] for more efficient tracking [3,4].

In the present case of cameras mounted on close-range scanning devices, highest accuracy in visual pose tracking is necessary as cameras feature small angular fields of view, which call for the concatenation of relative measurements (dead reckoning) so that errors readily accumulate. We propose **graph-based, nonlinear optimization** (keyframe-based bundle adjustment) on relative pose transformations and measurements, **parallel computing** of front-end, back-end and other sub-tasks, feature-based stereo vision, as well as **loop closure detection** for error compensation. Even in the case that everything else fails, **appearance-based recognition** of older features is provided so that pose tracking can be resumed. These contributions have been described in detail in Refs. [5,6].

Finally, since manual 3-D scanning requires visual feedback to the user, a **streaming surface reconstruction** method is presented that delivers realistic 3-D models *in-the-loop* during scanning as well as refined models promptly after loop-closing corrections.

We implement these methods on the DLR 3D-Modeler [7], creating the first 3-D scanner for close-range applications that localizes itself passively from its own images in realtime, at a high data rate. Systems of this type deliver more accurate results than depth sensors using coded infrared light (e.g., Kinect, Xtion) by an order of magnitude [8]. The DLR 3D-Modeler is a low-cost, hand-held device for accurate geometric and radiometric reconstruction of close-range objects in realtime that was originally tracked by robotic manipulators or external infrared light trackers (Fig. 1).

The remainder of this article is as follows: An extended survey on related 3-D modeling devices, their pose tracking techniques, and more specifically visual pose tracking is delivered in Section 2. In Section 3 we present the visual pose tracking algorithms implemented in the DLR 3D-Modeler. We validate the approach with experiments in Section 4 and supplementary videos.

2. State of the art

In this section we review 3-D modeling work with regard to their 3-D data registration concept—provided the system meets our requirements, i.e., is non-contact and light-weight. We focus on mature, commercial systems and only mention research work in the areas where commercial systems are missing. Lastly, we elaborate on the real-time variants of visual pose tracking for online 3-D data registration.

Fig. 1. The portable DLR 3D-Modeler used for cultural heritage preservation.

2.1. Data registration by scan alignment

Dense depth sensors that provide 2-D range images (i.e., 2.5-D images) yield rich surfaces that allow for data registration by 3-D matching, without the necessity for explicitly estimating sensor motion. This is not possible, however, in the case of 1-D range images (e.g., laser stripe triangulation).

3-D matching is computationally demanding because correspondence search is on higher dimensionality compared to traditional 2-D image registration. Additionally, data overlapping is required, which has to be detected in advance out of raw depth data and perhaps some motion priors. For these reasons, scan alignment is often being performed off-line, in an interactive way. The estimation involves an optimization in the form of the minimization of a distance metric between scans (e.g., ICP [9]). Different metrics and ICP modifications have been proposed for improved robustness against noise and efficiency [10]. With the recent advent of general-purpose computing on GPUs, real-time implementations of ICP have been presented (e.g., sequential multi-scale ICP on RGB-D data [11]). Other authors opt for bootstrapping ICP by feature-based visual pose tracking, see Ref. [12] and Section 2.4. Indeed, Coudrin et al. for the company Noomeo SAS use visual pose tracking for initial estimation for subsequent ICP optimization [10]. They are unable to use if for online data registration because they use densely projected patterns, which preclude concurrent visual tracking. They use interleaved stereo frames where the projected pattern is switched off, so that 3-D modeling and pose tracking are innerly desynchronized. In the end, half of the images serve 3-D modeling whereas the other serve as an initialization step for ICP.

2.2. Data registration by external pose tracking

3-D pointcloud registration is an over-determined problem with as few as 6 degrees of freedom (DoF). It is common practice to take data subsets to simplify the estimation problem. In addition, its convergence is subject to a high degree of unpredictability as it is strictly dependent on the particular surface geometry. We would benefit from a registration method that is independent of the 3-D data. It is well known that the sensor motion estimation problem (6 DoF) yields that same solution, although represented in the camera reference frame instead of in the object reference frame.

The use of traditional absolute positioning systems attached to a 3-D sensor is arguably the most straightforward approach for solving this problem. Due to their robustness and accuracy, the systems listed below became widespread and are the dominant (commercial) 3-D modeling devices in close-range:

- *External, optical tracking systems* are used by Northern Digital Inc., Metris NV, and Steinbichler Optotechnik GmbH. These systems detect and track artificial (e.g., infrared-reflecting) markers attached to the 3-D sensor. They seem convenient to hand-held operation due to the absence of a rigid contact to the tracking sensor. On second sight, however, the user feels strongly limited because of their small tolerance to sensor rotation owing to visibility constraints. Furthermore, since the spatial distribution of the markers is limited, the accuracy of orientation estimation is generally poor.
- *Passive arms* are used by FARO Technologies Inc., KREON Technologies, RSI GmbH, Metris NV, and ShapeGrabber Inc. Passive arms, or even robotic manipulators, are inconvenient for