Digital Signal Processing 35 (2014) 37–44

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Robust surface tracking in range image sequences

Farzad Husain [a,*], Babette Dellen [b], Carme Torras [a]

[a] Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028, Barcelona, Spain
[b] RheinAhrCampus der Hochschule Koblenz, Joseph-Rovan-Allee 2, 53424 Remagen, Germany

ARTICLE INFO

Article history:
Available online 18 September 2014

Keywords:
Range video
Surface fitting
Tracking
Segmentation

ABSTRACT

A novel robust method for surface tracking in range-image sequences is presented which combines a clustering method based on surface models with a particle-filter-based 2-D affine-motion estimator. Segmented regions obtained at previous time steps are used to create seed areas by comparing measured depth values with those obtained from surface-model fitting. The seed areas are further refined using a motion-probability region estimated by the particle-filter-based tracker through prediction of future states. This helps resolving ambiguities that arise when surfaces belonging to different objects are in physical contact with each other, for example during hand-object manipulations. Region growing allows recovering the complete segment area. The obtained segmented regions are then used to improve the predictions of the tracker for the next frame. The algorithm runs in quasi real-time and uses on-line learning, eliminating the need to have *a priori* knowledge about the surface being tracked. We apply the method to in-house depth videos acquired with both time-of-flight and structured-light sensors, demonstrating object tracking in real-world scenarios, and we compare the results with those of an ICP-based tracker.

1. Introduction

Tracking the pose of objects in image sequences is one of the most fundamental tasks in computer vision [1], and many works in the past focused on tracking in grayscale and color images. Tracking in range images is less explored, but due to the availability of low-cost depth cameras and their increasing importance in science and industry, such tracking approaches are of growing interest to the machine-vision as well as the robotics community. For example, tracking of object surfaces based on range data can be used to monitor and control the actions of a robotic arm during object-manipulation tasks [2]. Using depth information as the primary information source has the advantage that objects can be directly described by their geometric form, which is not affected by changes in the object's appearance in terms of color and texture, lighting conditions, shadowing or reflections. Furthermore, geometric features required for grasping are immediately available. Disadvantages of using depth cameras are their limited resolution, accuracy, and operating range. This poses special demands regarding robustness and adaptability for the algorithms dealing with this type of data.

Surface tracking in the domain of range image sequences has two main components: (i) extract the surfaces and establish the correspondence of the surfaces over the frames in the sequence of range images, and (ii) compute the motion transformation using these surface correspondences [3]. Both tasks are intertwined, as the correct extraction of surface patches helps finding correct correspondences, and vice versa. As long as surfaces are spatially disconnected, problem (i) can be more or less easily solved by clustering the 3D points based on their spatial proximity [2,4]. Problem (ii) can be solved by assuming 3D rigid-body motions between extracted point sets [3,5,6]. However, as soon as surfaces get in physical contact with each other, the problem becomes far more challenging, because in this case it is often impossible to distinguish between different objects based on depth differences alone. The situation becomes even more severe when both the manipulator and the manipulated surface undergo the same transformation at this time, e.g., during a hand-object manipulation. In this case, (i) and (ii) need to be solved conjointly, while taking the motion history of the objects into account.

In this work, we offer a solution to this problem by combining a recent clustering approach based on surface-model fitting [2] with a particle-filter-based affine-motion-estimation approach [7] with some modifications. Because we use a split-and-merge procedure for region growing which automatically adapts to the dynamic range data, predictions from the particle filter can be

* Corresponding author.
E-mail addresses: shusain@iri.upc.edu (F. Husain), dellen@hs-koblenz.de (B. Dellen), torras@iri.upc.edu (C. Torras).

http://dx.doi.org/10.1016/j.dsp.2014.09.011
1051-2004/

incorporated in a straightforward manner by refining seeding and, in consequence, the input to the tracker.

2. Related work

Object tracking has previously been performed mostly for color/gray-scale image sequences [1,7–10]. However, depth images pose different challenges to the tracking algorithm than color/gray-scale images.

Most methods for tracking in range images use *a priori* knowledge of 3D point correspondences and find the affine or rigid-body transformation on this basis [11]. These methods mainly work for sparse data sets, but are less useful when working with dense range images. Other techniques match surface patches instead, eliminating the need for finding exact point correspondences [3]. The range data is segmented into surface patches, then correspondences are established between patches of adjacent frames, and the motion transformation is estimated. However, such an approach is only effective if the initial (presumably correct) segmentation can be maintained over time. This is however not a trivial task, as small variations in the data and motions can change the segmentation drastically. To overcome this problem, a seeding and region growing technique for range-image sequences was proposed in [2,4]. Maintenance can be improved this way, but when two or more surfaces are in physical contact with each other, it remains difficult to determine the boundary between the surfaces in contact using depth differences alone [2].

To cope with the specific characteristics of range data, some existing approaches put limitations on the tracked surface by considering only articulated motion [12–14]. This simplifies the tracking problem but also limits the usability of the algorithm to particular scenarios. Robust tracking of human hands assuming articulated motion constrained by the 54-dimensional parameter space has been performed in [15]. In [16], object tracking using a depth camera was performed (for 3D object reconstruction), but here the robotic hand had to be separated from the range data before applying the algorithm.

Another option for 3D tracking is the Iterative Closest Point (ICP) algorithm. However, the basic ICP method [17] is a pairwise matching algorithm which does not take into account past measurements [18,6], hence the error starts to propagate. The ICP algorithm has been previously combined with Kalman filtering for object reconstruction [19]. However, in this case, the background was removed, leaving only the target object. In cluttered scenes, this approach may thus not be applicable. Point-to-point matching in 3D space requires a high accuracy in the estimation of the transformation matrix. This makes these approaches less suitable for our scenario because of the limited accuracy of the depth camera. Several variants of ICP which achieve better point set registration have been proposed, such as the expectation-maximization ICP [20] and softassign [21]. However, these variants have a higher computational cost and require specialized hardware such as GPUs to achieve real-time performance [22].

In this work, we combine seeding and growing of surfaces with particle filtering to overcome the aforementioned limitations. Our main contribution is a robust mechanism for identifying a set of points belonging to a target object that is being manipulated in 3-D space, regardless of its physical contact with other objects.

3. Method

Our tracker requires a range image as input at each time step. The range image along with the camera's intrinsic parameters is used to construct a 3-D point cloud. The algorithm for surface tracking consists of the following steps. Initially a set of non-overlapping geometric surface patches are obtained by clustering the 3-D points. Each cluster is modeled by a quadratic function and the surface that we want to track is identified manually (see Section 3.1). Segmented surfaces at step t are used to create seed regions in the next frame $t+1$ by comparing the predicted depth values (from quadratic surface models fitted to the segments) to the actual depth values (see Section 3.3). At the same time, a motion-probability region is found by the particle-filter-based tracker through the prediction of future states (see Section 3.2). The extracted motion-probability region is used to refine the seeding. Region growing allows reconstructing the segment at $t+1$. Based on this segmented area, the translation parameters of resampled states are re-estimated (see Section 3.4), which, provided the segmentation is correct, improves the predictions of the tracker for the next frame. The basic idea behind our approach is illustrated in Fig. 1.

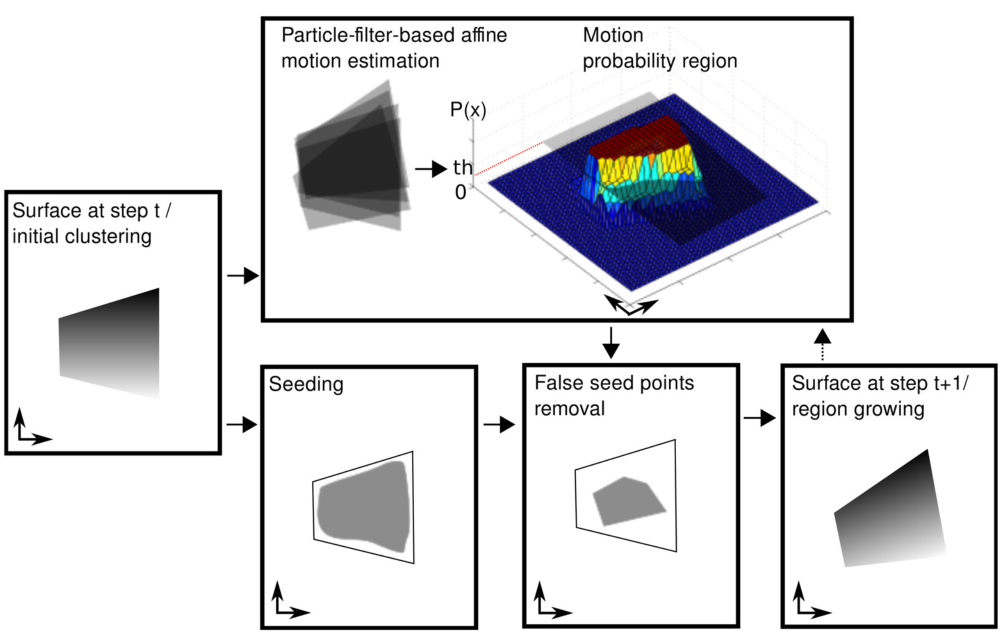

Fig. 1. Schematic illustrating the basic idea of our approach (for further explanations, see main text).