Pattern Recognition 47 (2014) 2962–2973

Contents lists available at ScienceDirect

Pattern Recognition

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

3D soft-tissue tracking using spatial-color joint probability distribution and thin-plate spline model

Bo Yang [a,b], Wai-Keung Wong [a,*], Chao Liu [c], Philippe Poignet [c]

[a] Institute of Textiles and Clothing, The Hong Kong Polytechic University, Hong Kong, China
[b] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China
[c] LIRMM, UMR 5506, CNRS-UMII, 34095 Montpellier, France

ARTICLE INFO

Article history:
Received 11 December 2012
Received in revised form
27 January 2014
Accepted 23 March 2014
Available online 4 April 2014

Keywords:
Robotic surgery
Visual tracking
Motion compensation
Stereo vision
Kernel function

ABSTRACT

Visual tracking techniques based on stereo endoscope are developed to measure tissue motion in robot-assisted minimally invasive surgery. However, accurate 3D tracking of tissue surfaces remains challenging due to complicated deformation, poor imaging conditions, specular reflections and other dynamic effects during surgery. This study employs a robust and efficient 3D tracking scheme with two independent recursive processes, namely kernel-based inter-frame motion estimation and model-based intra-frame 3D matching. In the first process, target region is represented in joint spatial-color space for robust estimation. By defining a probabilistic similarity measure, a mean-shift-based iterative algorithm is derived for location of the target region in a new image. In the second process, the thin-plate spline model is used to fit the 3D shape of tissue surfaces around the target region. An iterative algorithm based on an efficient second-order minimization technique is derived to compute optimal model parameters. The two processes can be computed in parallel. Their outputs are combined to recover 3D information about the target region. The performance of the proposed method is validated using phantom heart videos and in vivo videos acquired by the daVinci® surgical robotic platform and a synthesized data set with known ground truth.

1. Introduction

Robotic technologies are increasingly used to overcome the limitations of traditional minimally invasive surgery (MIS) tools and provide surgeons with improved visualization and enhanced dexterity. However, respiratory- and cardiac-induced tissue motion can impose significant challenges on delicate tasks such as vessel anastomosis. Even with mechanical stabilizers, residual tissue motion still needs to be manually canceled by surgeons, which considerably disturbs the execution of surgical procedures and results in longer surgery sessions and more surgical risks. Active motion compensation technology has been explored since 2000 [1–9]. The concept of heartbeat synchronization was first proposed by Nakamura et al. [1]. By tracking tissue motion and actively synchronizing surgical instruments with motion, it is possible to provide a virtually stable operating environment for surgeons. To achieve this, accurate estimation of tissue motion is needed. For other advanced MIS techniques, for example pre- and intra-operative surgical guidance [10] and dynamic active constraints [11], accurate estimation and tracking of tissue motion are also important prerequisites.

Various kinds of sensor systems have been employed to measure tissue motion, such as structured lighting [2,3], accelerometers [4], visual systems with artificial markers [5,6], ultrasound [7,8] and whisker sensors [9]. However, as indicated by Mountney et al. [11], these systems are not clinically popular as they require an additional instrument port in addition to the endoscope. From a practical point of view, passive visual tracking techniques based on natural structures of images are more appropriate for MIS. Gröger et al. [12] proposed a vision-based method by employing an affine motion model for tracking natural landmarks on the heart surface. It was subsequently improved by Ortmaier et al. [13] using an elegant prediction technique. Some classical methods for visual tracking, such as the color-based mean shift [14] and the efficient second-order minimization (ESM) [15], were introduced to tissue tracking [16,17]. However, previous studies were mainly limited to 2D issues until the advent of the stereo endoscope. The visual tracking technique with the stereo endoscope makes full 3D motion compensation possible. However, performing 3D tracking in MIS is challenging since endoscope images can be low in quality and disturbed by specular reflection, smooth or other dynamic effects.

* Corresponding author. Tel.: +852 27666471; fax: +852 27731432.
E-mail addresses: yangbo.sd@gmail.com (B. Yang), calvin.wong@polyu.edu.hk (W.-K. Wong), liu@lirmm.fr (C. Liu), poignet@lirmm.fr (P. Poignet).

http://dx.doi.org/10.1016/j.patcog.2014.03.020
0031-3203/

In this study, two basic problems in 3D tracking are inter-frame motion estimation and intra-frame 3D matching (see Fig. 1). To establish correspondences in the inter- and intra-frame, both feature-based and region-based methods are investigated. Feature-based methods represent track targets in feature space, whose performance is largely determined by reliability of feature descriptors. Maximally stable extremal regions and traditional gradient-based image features were employed by Stoyanov et al. [18] and later combined with a constrained geometrical surface model for robust tissue tracking [19]. The ID3 decision tree [20] and the anisotropic feature detector [21] were also introduced to select reliable descriptors. A real-time long-term tracking framework was developed by combining the STAR feature detector and binary robust independent elementary features for accurately tracking tissue in stereo-endoscopic scenes [22]. However, selecting a feature descriptor is context-specific and the appearance of tissue that varies greatly (from homogeneous to highly textured) may not contain special sets of features for tracking.

Region-based methods represent the target as a template using pixel intensity directly in image space. Deformable models are commonly used to connect pixels in the template with pixels in a new image. Both linear and nonlinear models have been employed for tissue tracking, such as affine models [12,13], B-spline [23], piecewise bilinear mapping [24], physically inspired models [25,26], image-constrained biomechanical models [27], and quasi-spherical triangle models [28]. In [29], an epipolar-based 3D tracking framework without explicit 3D reconstruction was proposed, which can be used for medical imaging. Computational analysis of deformation model was conducted based on interpolation methods to help create both a common tracking framework and a robust solution for heart motion compensation [30]. Since the whole pixel information is used, region-based methods can usually provide more accurate estimation than feature-based methods. However, in highly dynamic environments, accurate and stable inter-frame correspondences are difficult to establish because inter-frame tissue deformation is complicated and difficult to specify by an explicit model, and pixel intensities of the target region are susceptible to noise, motion blur and other dynamic effects in MIS. The inter-frame discrepancy makes similarity measures in image space error-prone, resulting in unstable tracking flow.

Richa et al. [31,32] have recently employed an extended thin-plate spline (TPS) model and a single error sum of squares (SSE) measure to perform both the inter-frame and the intra-frame matching tasks. The TPS is a physical model with minimal bending energy for describing the 3D shape of tissue surfaces, and the pixel-based SSE measure is efficient for intra-frame 3D matching due to good consistency between right and left images of the stereo pair. However, for inter-frame motion estimation, the SSE-measured method suffers the same problems as the aforementioned methods. With an increase in inter-frame discrepancy over time, independent pixel information may not be sufficient to distinguish a region from its surroundings.

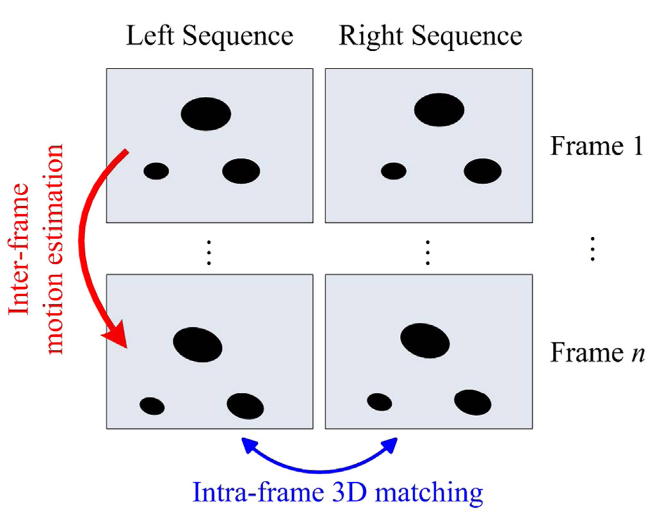

Fig. 1. Inter-frame motion estimation and intra-frame 3D matching for 3D tracking.

In this study, a robust and efficient tracking scheme is proposed. By considering different characteristics of inter-frame images and intra-frame stereo-pair images, two independent recursive processes are employed for inter-frame estimation and intra-frame matching, respectively. In order to be robust to inter-frame discrepancy, the target region is represented in joint spatial-color space with multivariate kernel density estimation. A probabilistic similarity measure is employed and a mean-shift-based iterative algorithm is derived to locate the target region in new images. To recover 3D information, stereo-pair images are matched directly in image space using the TPS model due to good consistency between stereo images. Relative pixel coordinates are introduced to the TPS so that 3D matching for consecutive frames can be performed efficiently using a constant model matrix. An ESM-based iterative algorithm is derived for computing optimal model parameters. For efficient 3D tracking, a parallel scheme is designed based on the two processes. Experiments on various data sets and comparisons with existing methods are conducted to justify the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section 2 proposes the kernel-based inter-frame motion algorithm. The model-based intra-frame matching is discussed in Section 3. Section 4 presents the parallel 3D tracking scheme. In Section 5, experimental results and comparisons are provided. Conclusions are given in Section 6.

2. Inter-frame motion estimation

The purpose of tissue tracking is to measure the motion of a point of interest (POI) in 3D space. The region of interest (ROI) is the support region around the POI, which makes accurate POI tracking possible. In this section, the POI is estimated in a 2D (left or right) image plane. Since pixel-based target representations and similarity measures in image space are sensitive to inter-frame discrepancy, the ROI is represented and tracked in joint spatial-color space. The joint spatial-color probability distribution is first reviewed. A probabilistic similarity measure is then introduced, which can be considered as a generalization of the classical SSE measure and the color-histogram measure. A mean-shift-based optimization algorithm is derived to iteratively compute the position of the POI in new frames.

2.1. Joint distribution representation

The pixel in an image I can be considered as a random variable in the joint spatial-color space, $\mathbf{s}=[\mathbf{m},\mathbf{c}]^T$, where $\mathbf{m}=[u,v]^T$ denotes 2D pixel coordinates and $\mathbf{c}=I(\mathbf{m})$ is a 3×1 RGB color vector. Let Λ denote the initial ROI, a rectangular region centered at $\overline{\mathbf{m}}$ (the POI) in the reference frame (normally the first processed frame). Given the set $\mathbb{S}=\{\mathbf{m}_i,\mathbf{c}_i\}_{i=1}^N$ where N is the number of pixels inside Λ, the probability distribution $p(\mathbf{m},\mathbf{c})$ in the joint space can be estimated using the multivariate kernel density estimation [33,34]:

$$p(\mathbf{m},\mathbf{c})=\frac{1}{N}\sum_{i=1}^{N}K_{h_s}(\mathbf{m}-\mathbf{m}_i)K_{h_c}(\mathbf{c}-\mathbf{c}_i),\ \text{with} \tag{1}$$

$$K_h(\mathbf{x})=\exp\left(-\frac{\|\mathbf{x}\|^2}{2h^2}\right) \tag{2}$$