Robotics and Autonomous Systems 97 (2017) 182–192

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Needle in a haystack: Interactive surgical instrument recognition through perception and manipulation☆

Tian Zhou, Juan P. Wachs *

School of Industrial Engineering at Purdue University, West Lafayette, IN 47906, USA

HIGHLIGHTS

- A machine vision to pick cluttered surgical instruments for a robotic scrub nurse.
- The recognition approach involves visual perception and dexterous manipulation.
- A force-based reactive grasping protocol is proposed.
- A new segmentation algorithm based on parts-based detection is presented.

ARTICLE INFO

Article history:
Received 7 September 2016
Received in revised form 27 July 2017
Accepted 25 August 2017
Available online 7 September 2017

ABSTRACT

This paper presents a solution to the challenge of accurate surgical instrument recognition by a Robotic Scrub Nurse (RSN) in the Operating Room (OR). Surgical instruments, placed on the surgical mayo tray, are often cluttered, occluded and display specular light which poses a challenge for conventional recognition algorithms. To tackle this problem we resort to a hybrid computer vision and robotic manipulation combined strategy. The instruments are first segmented and pose is estimated, then the RSN system picks up the unknown instruments and presents them to the optical sensor in the determined pose. Last, the instruments are recognized and delivered. Experiments were conducted to evaluate the performance of the proposed segmentation, grasping and recognition algorithms, respectively. The proposed patch-based segmentation algorithm can achieve an F-score of 0.90. The proposed force-based grasping protocol can achieve an average picking success rate of 92% with various instrument layouts, and the proposed attention-based instrument recognition module can reach a recognition accuracy of 95.6%. Experimental results indicate the applicability and effectiveness of a RSN to perform accurate and robust surgical instrument recognition.

1. Introduction

American hospitals are facing critical problems of lacking nurses. According to a recent report [1], there will be a shortage of 260,000 registered nurses by 2025 in the USA. Such shortage could lead to an increase in mortality. For example, it was reported that patient mortality risk is 6% higher in hospitals understaffed with nurses compared to units fully staffed [2]. One solution to the nurse shortage problem is to develop systems which can take over the highly mechanical, mundane and repetitive tasks from the nurses, so as to allow them to focus more on complex tasks. Among all the tasks that the nurses are responsible for, the surgical instrument preparation and delivery task is one of the most important and dominant cases, which the proposed Robotic Scrub Nurse (RSN) system is trying to address and take over.

To build a robust RSN system, accurate instrument recognition is a critical component. Although humans can perform such tasks elegantly with the help of both vision and tactile information, it is quite a challenging task for robots to accurately recognize them. The instruments are often clustered together on the mayo stand and have a reflective and uniform appearance in both shapes and colors, as shown in Fig. 1. To enable a RSN system to serve robustly in the Operating Room (OR), it is critical for the robot to recognize the instrument accurately, localize them precisely and manipulate them robustly. The aim of this paper is to propose one such system which meets the aforementioned requirements.

☆ Research supported by the NPRP award (NPRP 6-449-2-181) from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

* Corresponding author.
E-mail addresses: zhou338@purdue.edu (T. Zhou), jpwachs@purdue.edu (J.P. Wachs).

http://dx.doi.org/10.1016/j.robot.2017.08.013

Fig. 1. A realistic mayo tray configuration.

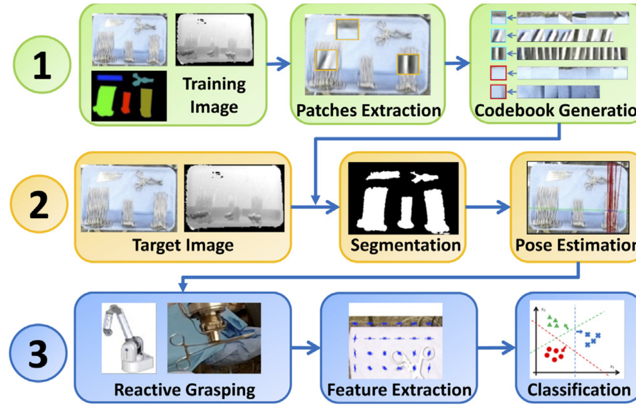

Fig. 2. System Architecture.

2. Related work

Robotic systems have been brought into the OR in mainly three different forms [3], (1) handheld robotic tools; (2) teleoperated surgical systems and (3) autonomous surgical assistants. The first two cases focus on designing robots to increase surgeon's sensorimotor capabilities. Such systems include the minimally invasive telesurgery system da Vinci [4], the steady-hand robotic system for microsurgical augmentation [5] and the touchless telerobotic surgery system which allows free hand gesture interactions [6]. The third category refers to robots assisting the surgeons and nurses without directly touching the patient. One specific type of robotic surgical assistant is the RSN system which manages, delivers and retrieves surgical instruments for surgeons upon request. Such RSN system can free the nurses to perform multiple tasks concurrently [7] and has the potential to reduce communication errors [8]. Many of the previous efforts in RSN development have been spent on the human–robot interaction part, where gestures [9,10], speech [11], haptics [12], EEG/EMG [13] and body-worn sensors [14] were used to enable communications between surgeons and RSN systems. However, there is a lack of attention to the problem of instrument recognition and manipulation in realistic/uncontrolled settings. For example, non-overlapping instrument on a specially-designed mayo platform are addressed [15]. Alternatively, the instrument locations were recorded ahead of time and their identities were recognized through infrared labels [12]. This work tried to bridge the gap of recognizing surgical instruments in an uncontrolled surgical setting without the usage of additional labels.

The proposed recognition algorithm consists of instrument segmentation, reactive grasping and instrument recognition. These problems are at the core of this paper and the relevant literature are discussed in the following.

Image segmentation often involves dividing the entire image into smaller coherent groups for further processing. The three major approaches are edge-based, region-based and threshold-based methods. The edge-based approach first finds edges on the image using various filters (e.g., wavelet transformation [16], Gabor filter [17]), and then links the adjacent edges to form a contour (e.g., K-means [18], Fuzzy C-means [19]). However, they are sensitive to noise and do not work well on images with smooth transitions and low contrast. On the other hand, the region based approach consider the grayscale similarities in neighboring pixels to merge and split them into sub-regions, e.g., Watershed [20], Chan–Vese model [21], superpixel [22], GrabCut [23]. Some of the region-based algorithms require initial seed region to start the segmentation (by human or using some heuristic). The last group, threshold-based method, compares the image intensity levels with a threshold, which are found automatically using Otsu's [24], genetic algorithms [25] or particle swarm optimization [26], to mention a few. More recently, Convolutional Neural Network (CNN) algorithms were used to achieve both segmentation and recognition with satisfactory accuracies [27,28]. However, these algorithms have a large memory foot-print and an ambitious hardware requirement for usage. The proposed patch-based segmentation algorithm solves both segmentation and recognition at the same time, and does not require such hardware.

The instrument recognition algorithm follows an interactive recognition strategy, where the object-of-interest is grasped, manipulated and observed actively by a robot to determine its identity. Lyubova et al. proposed a framework where a humanoid robot is able to recognize unknown objects through interaction and observation [29]. Such procedure mimics the process of human learning, and increases the recognition accuracy by cumulating the confidences from multiple modalities (i.e., vision, tactile and affordance [30]). However, to use such approach, the robot is required to be able to grasp unknown objects successfully, which on the other hand usually requires accurate object identity and properties, such as their 3D models [31] or physical properties [32]. This "chicken-and-egg" problem can be solved through reactive grasping protocols, which do not require full object models, and are at the core of our work. Some reactive grasping algorithms leverage tactile feedback [33] and optical proximity sensors [34] to grasp the unknown object robustly. The proposed force-based reactive grasping protocol also works without full object models.

This paper makes the following contributions: (1) a segmentation algorithm based on visual codebook generation and weighted histogram backprojection; (2) a force-based reactive grasping protocol to enable reliable instrument grasping; (3) a surgical instrument recognition algorithm. All these components are key for the successful implementation of a RSN.

3. System architecture

Recognition and grasping of surgical instruments from mayo trays in a clinical setting is challenging due to: (a) the instruments are packed together leading to occlusion, discontinuities and clutter; (b) salient points are difficult to find on metallic instruments due to their uniform and reflective composition.

We tackled this challenge through an active recognition process, which involves manipulation and then recognition. The system architecture is shown in Fig. 2. Firstly, a codebook of local appearances is created in the training phase (step #1). During usage, the packs of instruments are segmented and their poses