2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

A Vision-Guided Robot Manipulator for Surgical Instrument Singulation in a Cluttered Environment

Yi Xu, Xianqiao Tong, Ying Mao, Weston B. Griffin, Balajee Kannan, and Lynn A. DeRose

***Abstract*— The logistics of counting, sorting, sterilizing, and transporting surgical instruments is labor and capital intensive. Furthermore, infection due to improper sterilization is a critical safety hazard. To address these problems, we have developed a unique robotic manipulation system that is capable of accurately singulating surgical instruments in a cluttered environment. Our solution is comprised of two parts. First, we use a single-view vision algorithm for identifying surgical instruments from a pile and estimating their poses. Occlusion reasoning is performed to determine the next instrument to grip using a contrast invariant feature descriptor. Second, we design a compliant electromagnetic gripper that is capable of picking up the identified surgical instrument based on its estimated pose. We validate our solution through instrument singulation experiments demonstrating identification, localization accuracy, and robustness of occlusion reasoning as well as the flexibility of the electromagnetic gripper.**

I. Introduction

The perioperative setting is considered the most resource intensive section of the hospital. Its performance has a significant impact on patient safety and operating budget. Currently, processing hundreds of types of surgical instruments with very similar characteristics requires an extensive learning curve for hospital employees and perfect human diligence. Automating the process has the potential to significantly address current safety and efficiency concerns. During the process of automated sorting, a static robot arm picks up each instrument and subsequently sorts the instruments into different bins or stacks. An end-effector must be attached to the robotic arm to enable handling of a varied suite of instruments in highly cluttered bins.

Existing approaches to automating the sorting process are expensive and are limited in their capabilities. The state-of-the-art solution in this domain, the RST PenelopeCS™ is designed to automate several key functions for the clean side of the sterile supply [1]. With this system, a human operator first separates the surgical instruments from a container, and places them on a conveyor belt one by one. Then a standard robotic arm fitted with a magnetic gripper picks up the single instrument from the belt. A machine vision system or a barcode scanner is used to identify the instrument and sort them into stacks of similar instruments. While effective, this solution requires additional infrastructure as well as human operation at critical junctures. Consequently, such solutions are limited when handling instrument manipulation in an unstructured environment where surgical instruments are cluttered in the container. There are several challenges in designing such a robot. First, the surgical instruments often have very similar characteristics. This makes it difficult for vision-based algorithms to recognize them from an unordered pile. Second, instruments are made of shiny metal. Optical effects such as specularities and interreflections pose problems for many computer-vision based pose estimation algorithms, including multi-view stereo, 3D scanning, etc. Without full 6 degrees-of-freedom (DOF) pose, many standard grippers will have trouble executing the grip. Our solution addresses all these challenges effectively.

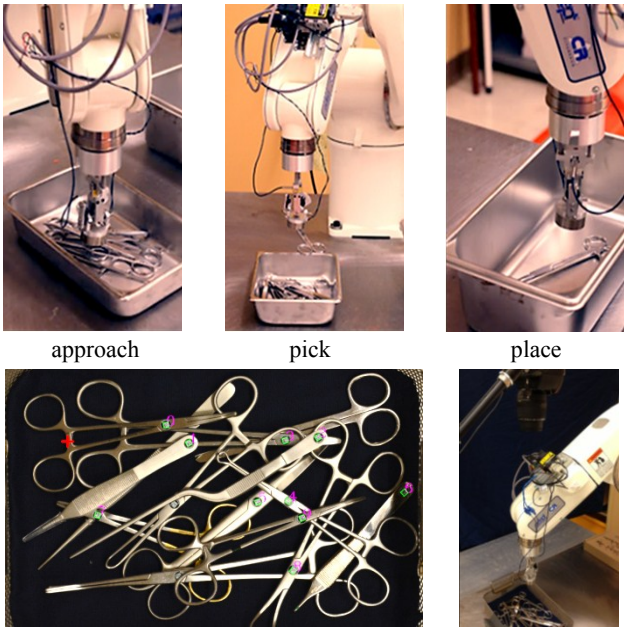


Figure 1. (Top) Surgical instrument pick-and-place using our robotic manipulation system. (Bottom left) A picture captured by the camera looking at the container. Our algorithm computes a 2D reference point as grip location (highlighted using red "+") for an instrument that is on top of the pile. (Bottom right) A picture of our system setup.

Our approach to automated sorting has two key elements: first, a robust vision algorithm that identifies the instruments, infers the occlusion relationships among the instruments, and provides visual guidance for the robot manipulator; second, a unique end-effector design, which is attached to a six-axis industrial robotic arm and can execute precise instrument gripping in cluttered environment with only a 2D reference point as picking location. This flexibility of the end-effector is important because determining a weak 4-DOF pose in 2D space is easier and potentially faster than computing an accurate full 6-DOF pose due to the optically-challenging nature of the surgical instruments. To our knowledge this is

Yi Xu is with GE Global Research, Niskayuna, NY 12309 USA (phone: +1-518-387-5946; fax: +1-518-387-5589; e-mail: xuyi@ge.com).

Xianqiao Tong is with Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA. The work was done when he was an R&D intern with GE Global Research (e-mail: txq1986@vt.edu).

Ying Mao, Weston B. Griffin, Balajee Kannan, and Lynn A. DeRose are with GE Global Research, Niskayuna, NY 12309 USA (e-mails: {yingmao|griffin|Balajee.Kannan|derose}@ge.com).

the first instance of an automated sorting solution that is robust to handling a varied instrument suite. Figure 1 demonstrates our robotic system for surgical instrument manipulation. Our contributions include:

- An integrated vision-guided robot manipulator for singulating individual surgical instrument from a cluttered environment.
- A computer vision algorithm that identifies the instruments, estimates 4-DOF object poses, and determines the order of objects from a pile.
- A custom electromagnetic gripper with multi-axis compliance that grips surgical instruments with only a 2D location as reference.

The rest of the paper is organized as follows. Section II outlines the related work in the area of vision-based bin-picking in unstructured environments as well as the current state-of-the-art in end-effector design. In Section III, we describe the details of our vision-based algorithm that provides visual guidance for the robot manipulator; while in Section IV we detail our electromagnetic end-effector design. The results from our experiments are summarized in Section V. Finally, Section VI provides a conclusion and outlines a brief plan for future research.

II. Related Work

A. Vision-Guided Picking

Picking individual items from an unordered pile in a container or in unstructured environments has been a popular research topic for several decades [2]. Some of the recent developments rely on active sensing. For example, Kinect sensor has been used to acquire a depth map of the scene [3][4]. Then, known 3D object models are matched to the acquired point clouds. Choi et al. [5] use a structured-light 3D sensor to capture 3D models of objects in a box. A Hough-voting approach based on oriented surface points is used to estimate pose of the objects. A 3D laser scanner is used in PR2 for mobile manipulation in an unstructured environment [6]. These 3D sensors will have difficulty capturing specular objects such as surgical instruments.

Some previous work has aimed to provide a solution for bin-picking shiny objects in industrial settings. Shroff et al. [7] propose a system that extracts high curvature regions on specular objects using a multi-flash camera. Multi-view triangulation on these features is then used to obtain the object pose. Rodrigues et al. [8] build an imaging system with a single camera and multiple light sources. A random fern classifier is trained to map the appearance of small patches into poses. Liu et al. [9] use a multi-flash camera to obtain good depth edge information and use fast directional chamfer matching to match templates onto the input image for robotic bin-picking. Pretto et al. [10] use a single camera and a large diffuse light source mounted over the container. Their approach estimates a 6-DOF pose for planar objects based on a candidate selection and refinement scheme. In their experiments, each container only contains one object type. Both this method and the chamfer matching based methods rely on minimizing a cost function over a parameter space of transformations for each template. Computation time scales linearly with the number of templates. While for our approach, the computation time depends on the actual number of objects and their image-space intersections instead of number of templates. This makes our method suitable for handling large variety of surgical instruments. Large number of templates also increases the proportion of false correspondences for template matching, especially when the surgical instruments are very similar to each other. In addition, our method uses one single DSLR camera without the need of multiple lights and multi-views. Thus, it is cheaper and easier to implement. Instead of 6-DOF pose, weak 4-DOF pose in 2D space is estimated to guide the gripper.

Data-driven approaches (e.g., [8]) are also popular among robotics researchers. Collet et al. [11] develop a multi-view approach that can recognize all objects in the scene and estimate their full poses. Their approach relies on learned models for the objects using SIFT descriptor. Due to strong similarities among surgical instruments, such a data-driven method will be difficult to carry out.

Occlusion reasoning has also gained attention in the vision community. This problem can be solved by modeling local inconsistency for the occluders [12]. However, for surgical instruments, occluders will have similar appearance as those being occluded. Alternative approaches focus on learning the structure of occlusions from data [13]. This method could potentially work for our application but requires a training stage. Hsiao and Hebert [14] propose a multi-camera approach to model the interactions among 3D objects. Our approach compares the single input image with different hypotheses generated based on all possible occlusion configurations. We use a contrast invariant feature descriptor that is robust to the challenging optical effects.

B. End-Effector Design

Pick-and-place robots are a mature class of robots that have increasingly become a critical functional component in the manufacturing and healthcare domains. Robotic solutions in these domain range from DaVinci surgical robot to the low-cost Baxter robot for close operations with humans. Despite the significant advances in the field, a key open technical challenge in the domain is the need for low-cost, adaptive end-effectors that can be used for precision operations. Commonly-used and commercially-available robot grippers include vacuum grippers, finger grippers and magnetic grippers. Vacuum grippers are widely used in industry for its simplicity and gentle manipulation. However, they have difficulty in handling irregularly-shaped objects with uneven surfaces; thus inappropriate for surgical instruments. Finger-type grippers have high precision and repeatability. Most of them are custom made and tailored to specific applications. Recent advances in dexterous manipulation have made it possible to grip a large variety of objects with just a single gripper design [15]. However, handling surgical instruments in cluttered environments is still very challenging. Sophisticated vision guidance with full 6-DOF pose estimation is needed to ensure a firm grip. Collision detection and avoidance is also required.

Magnetic grippers can generate enormous gripping force in a very compact form factor and can grip objects with irregular shapes. Because most metallic surgical instruments that require sterilization are ferromagnetic [16],

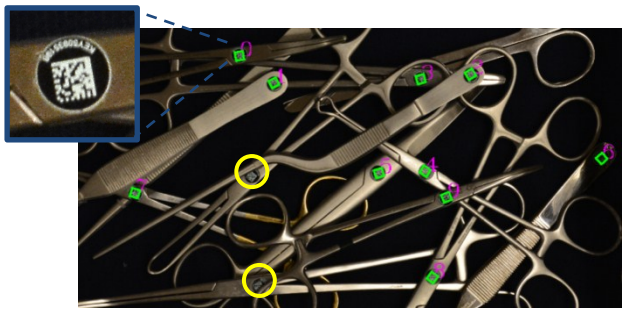


Figure 2. Decoded barcodes are highlighted using green bounding boxes. There are a couple of missed detections due to occlusion (yellow circles). (Inset) Zoom-in view of the barcode.

PenelopeCS™ has used a magnetic gripper for surgical instrument manipulation [1]. Their gripper has an electromagnet mounted on a 1-DOF spring-loaded base to allow for firm contact between the electromagnet and the surgical instrument. Such a design works well for instruments with known surface orientation (i.e., single object lying on a flat surface). However, system utility will be limited when used with instruments in a pile since 6-DOF pose estimation is required for obtaining contact surface orientation. To address this issue, we designed an electromagnetic gripper with 3-DOF compliance, which allows the electromagnet to passively reorient and conform to the instrument surface.

III. Vision-Based Instrument Localization

Typically, a sterile processing unit has a dirty side and a clean side. On the dirty side, instruments are washed and disinfected. Debris and biological material are removed from the instruments. On the clean side, instruments are counted, packed, and sterilized. Our system is designed for clean side operation. It first identifies the surgical instruments within the container for counting purpose. Then, it performs an occlusion reasoning step that determines what instruments are on top of the pile and not occluded by others. These instruments are candidates for gripping and singulation. Finally, a grip point is determined and communicated to the robot manipulator. The robot picks up the surgical instrument, and places it in an appropriate location for the purpose of sorting. This process is repeated again until all objects are removed from the container. To achieve these goals, we use a high resolution DSLR camera that is mounted over the robot manipulator. The camera looks downward at a pre-defined region within the robot's work space.

A. Surgical Instrument Identification

In recent years, automated tracking of surgical instruments has been gaining popularity among healthcare organizations [17][18]. Individual instrument tracking beyond tray level improves infection control and provides a mechanism for process improvement and root cause analysis. Typically, each surgical instrument is equipped with a 2D data matrix barcode encoding a unique ID for the tool. These barcodes are small—ranging from 1/8 to 1/4 inch in diameter, making them suitable for tracking objects with small flat surfaces such as surgical instruments.

Our system uses KeyDot [17] for instrument identification. A commercial off-the-shelf barcode reading module [19] is used to locate and read the IDs of all the visible barcodes from high-resolution images. To ensure

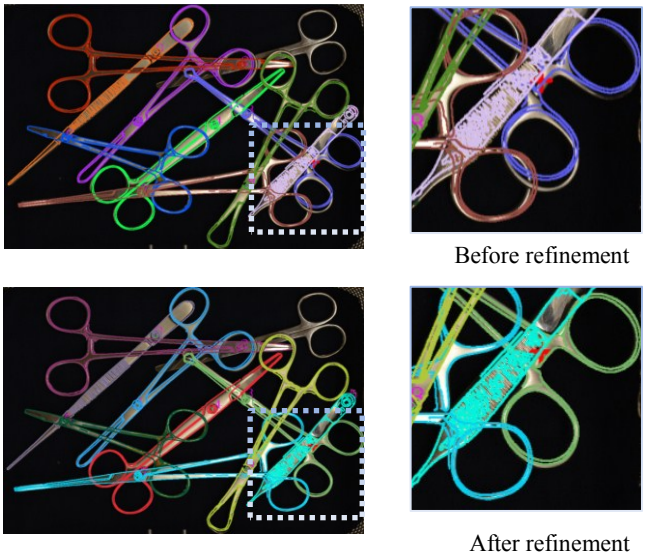


Figure 3. (Top) Pose estimation using the four corners of the data matrices from both template and input image of the container. (Bottom) Pose estimation after non-linear refinement. We superimpose transformed template edge maps onto the input image with random color. Zoom-in views show the better alignment after pose refinement.

identification, we place two barcodes on each side of an instrument. We assume flat objects, such as surgical instruments, only have two possible stable placements in the container. For non-flat instrument (e.g., forceps), a cap is used to close the tips, reducing the potential for alternate stable orientations that could limit barcode visibility. Since our system operates on the clean side of sterile processing, the instruments are free of debris and biological material. Barcodes will not be contaminated. Figure 2 shows an image of the instruments in a container and detected barcodes highlighted in green.

B. Localization and Pose Estimation

Given the instruments' IDs in the container, our system estimates 4-DOF pose (i.e., location, orientation, and scale) for each instrument. We achieve this by matching template of each instrument to its pose in the image of the container. A library is populated with templates in a pre-processing step. Each template is an image of an instrument captured with black background. We segment the template by thresholding and removing small connected components. The foreground segmentation is also stored in the library as a mask.

For each detected barcode within the container, we retrieve its counterpart template from the library. The barcode reading module not only reads the ID, but also detects the four corner points of the data matrix. By using the four corners on both the input container image and the template image, we compute an initial affine transformation that brings the template into alignment with the instrument in the container.

The estimated affine transformation is subsequently refined using a non-linear optimization. Because of the similar appearance of surgical instruments and occlusions within the scene, a typical cost function that minimizes the distance between re-projected and original point features is not feasible. Instead, we use a cost function that minimizes the re-projection error of edges. We first detect edges in both

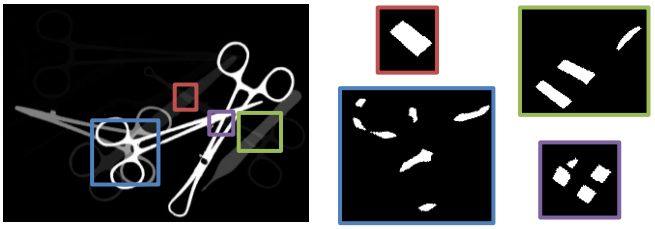


Figure 4. (Left) An occupancy map of the instruments. Instruments that are assigned to a lower bit are expected to have lower intensity. (Right) Several dilated binary masks showing pair-wise intersection regions between instruments.

template image and container image. Then, we compute the distance transform on the edge map of the container image. In this way, the distance between the transformed template edges and edges in the container image can be approximated quickly by summing up pixels at transformed template locations in the distance transform. We use downhill simplex to perform the non-linear optimization. Figure 3 shows a visualization of the instrument localization and pose estimation step. We superimpose the edges of transformed templates onto the input image to show the effectiveness of the alignment.

It is noteworthy that although projective transformation is more accurate for pose estimation of tilted surgical instruments, we use affine transform because it has fewer parameters (6 vs. 8), and because of the ability to obtain good initial guess from the data matrix corners. As we will discuss later, due to our robust occlusion reasoning algorithm and our compliant end- effector design, perfect pose estimation is not essential for the success of the gripping task.

C. Occlusion Reasoning

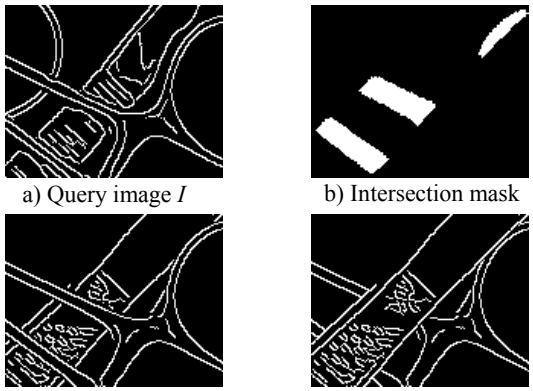
Given the poses of instruments whose barcodes are visible, our system then infers the occlusion relationship between each pair of intersecting instruments. This determines which instruments are not occluded by others and thus possible for the gripper to pick up next.

1) Finding the Intersections

Our system first computes an *occupancy map* that models the intersections between all instruments. The occupancy map is a single channel image. Each bit of a pixel is assigned to one instrument. We first transform each template and its foreground binary mask using the computed affine transformation. We then store the transformed binary mask in the designated bit in the occupancy map. In this way, the intersecting region between two instruments can be easily determined by finding the occupancy map pixels that have the two corresponding bits on. Figure 4 shows an occupancy map and the intersection region between a few pairs of instruments.

2) Inferring the Occlusions

For every pair of intersecting surgical instruments A and B, there are only two hypothetical occlusion relationships: A occludes B or B occludes A. For each pair, we synthesize two images each of which corresponds to one of the two hypotheses (H_1 and H_2). The images are synthesized by transforming the templates to their poses in the input image and rendering them in two different sequential orders. The occlusion relationship can then be inferred by comparing the input image I against the two hypothesis images. H_1 and H_2 only differ at the intersecting regions and are identical for the rest of the pixels. We use the intersecting region computed from the occupancy map as a binary mask and only compare I against H_1 and H_2 within the masked region. We dilate the mask by a small amount to account for inaccuracy in the estimated instrument pose and to ensure all intersection regions are included in the binary mask.



a) Query image I b) Intersection mask

c) H_1 d) H_2

Figure 5. a) A query image I generated by Canny edge detection on the input image and b) the associated intersection mask. c-d) The two synthesized hypotheses

To compare images, we use a descriptor called Edge Orientation Histograms (EOH) [20]. EOH descriptors are similar to the widely used Histograms of Orientated Gradients (HOG) [21]. EOHs are contrast invariant and only use edges instead of appearance information. They can handle large appearance changes due to varying lighting conditions, specularities, and interreflections among the surgical instruments. To reduce noise and focus on the more important contour edges, the input images are first blurred with a Gaussian kernel. We compute masked EOH (denoted as *meoh)* descriptors for image I, H_1, and H_2 and then compute Eculidian distances between the histograms. The hypothesis H with the smaller histogram distance to the input image I is selected:

$$\hat{H} = \arg\min_{H_i} \left\| meoh(I) - meoh(H_i) \right\|_2, \text{where } i = 1,2 \qquad (1)$$

Figure 5 shows a query image and two hypotheses. Notice that edges on the forceps (the bottom object) are different in the query image I and hypothesis images H_1 and H_2, because the templates and actual input image are captured at different lighting conditions. Due to strong presence of edges from the scissors (top object), our method is still able to select the correct hypothesis.

3) N-Instruments Intersection

In the case of more than two instruments intersecting at the same region, our method can still correctly predict the one that is on top of the pile. For example, if A occludes B and C, due to strong presence of A's edges in the query image I, the algorithm will predict A occludes B and C in two separate evaluations. The relationship between B and C is not important because we only look for one instrument that is on top. Once A has been picked up, the relationship between B and C will be determined again in a later iteration. In other words, we are not trying to sort the entire set in one step. Because the pile of instruments may shift during each

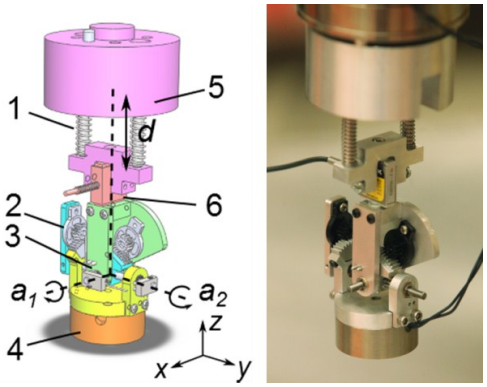


Figure 6. Design and prototype of the electromagnetic gripper. 1: Spring for the *z*-axis compliance. 2: Rotary damper. 3: Torsion spring. 4: Electromagnet. 5: Adapter for Adept robot arm. 6: Load cell.

gripping, we need to capture and process the scene before each grasping.

D. Picking Order Determination

We assume in most cases, there is no occlusion cycle among instruments (e.g., *A* occludes *B*, *B* occludes *C*, and *C* occludes *A*). This is a reasonable assumption considering all instruments are rigid, flat, and with tip closed. In case of instruments with occluded barcodes, because they will unlikely occlude the topmost instrument, our algorithm will still correctly predict the top one to pick.

Once all the occlusion relationships are determined, our algorithm finds the non-occluded ones. These can be instruments that do not intersect with others or that are on top of others. Our algorithm randomly selects one for the robot manipulator. The picking location for each surgical instrument is determined empirically by a human operator in the template creation stage. Since our camera captures a top-down view of the container, the image plane is parallel to the robot's working surface. Thus, the picking location in the image space can be easily translated into the robot's coordinate using a linear transformation. Because of our compliant end-effector design, imperfection in this transformation does not affect gripping accuracy much. The picking height (normal to the image plane) is approximated based on known information for the given container and table surface heights.

IV. End-Effector Design

The presented electromagnetic gripper is designed to grip a surgical instrument in a cluttered environment given a 2D reference point as the picking location. An annotated illustration and photo of the gripper is shown in Figure 6.

The electromagnetic gripper has three passive joints: a prismatic joint along the *z*-axis and two revolute joints along the *x* and *y* axes. Each joint has a coil spring attached to the shaft; making these joints compliant. A rotary damper (ACE Controls RTG2) is attached to each of the revolute joint shafts to prevent instrument oscillation after picking up and during transport. A load cell (Futek LSB200, 5lbs) is used to measure the contact force between the electromagnet and the instrument. We use a threshold force F_t (e.g., 7N) to determine when the electromagnet is in proper contact with an instrument (i.e., no gap between the electromagnet and instrument). The load cell also measures the weight of an instrument which can be used to detect gripping errors (e.g., multiple instrument gripping, loss of instrument, incorrect instrument, etc.). An electromagnet (APW EM137S) is attached at the bottom of the gripper. To reduce potential adherence between target instrument and adjacent instruments, the current of the electromagnet is modulated by a servomotor drive (A.M.C. AZBH12A8B) to generate a gripping force just enough to pick up the target instrument. We empirically determine a picking force for each instrument and store them in a lookup table indexed by the instrument ID, which is provided by the vision system in real time. The Adept robot arm and the electromagnetic gripper are controlled by an Adept SmartController. The load cell and the servomotor drive are interfaced with the robot controller through a Wago DeviceNet analog I/O module.

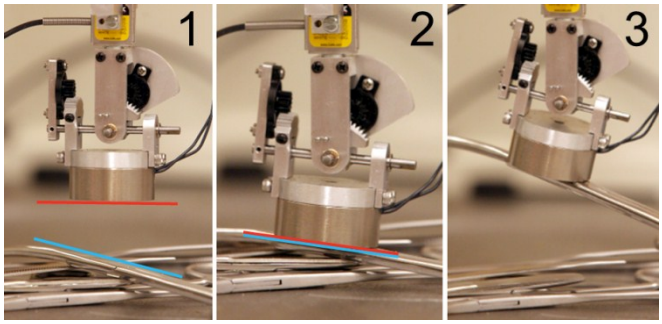


Figure 7. Workflow of instrument gripping using the electromagnetic gripper. The red/blue lines approximate the surface orientation of electromagnet and instrument before and during contact.

An illustration of instrument pickup workflow is shown in Figure 7. Given a 2D gripping location of a target instrument, a pick-up maneuver is completed with the following three steps:

1. The robot arm moves the gripper to the gripping location and stops at a distance *h* above the instrument. Distance *h* is experimentally determined (e.g., 8cm) to avoid any potential collision during this step.
2. The robot arm approaches the target along the *z*-axis. When the electromagnet comes in contact with the target, it re-orients itself to align with the instrument surface, regardless of their initial relative orientations. The robot controller constantly monitors the contact force until F_t is reached; indicating full contact with the target.
3. The electromagnet is energized to pick up the instrument. The current is adjusted to the pre-calibrated value. The robot arm moves the electromagnetic gripper to lift the instrument and removes it from the container.

In case where the manipulator is not able to grip an instrument successfully, the system will move on to the next available instrument. When the scene is further de-cluttered, the system will eventually attempt to grip the failed object again. In the rare situation where all instruments are occluded (i.e., an occlusion cycle is presence or a detection error in the vision algorithm), the algorithm selects the instrument that has the smallest number of occlusions and uses occlusion size as a tiebreaker. Because of the compliant design, gripping an instrument in such a situation is in general not a problem. After removing an instrument from the container, system

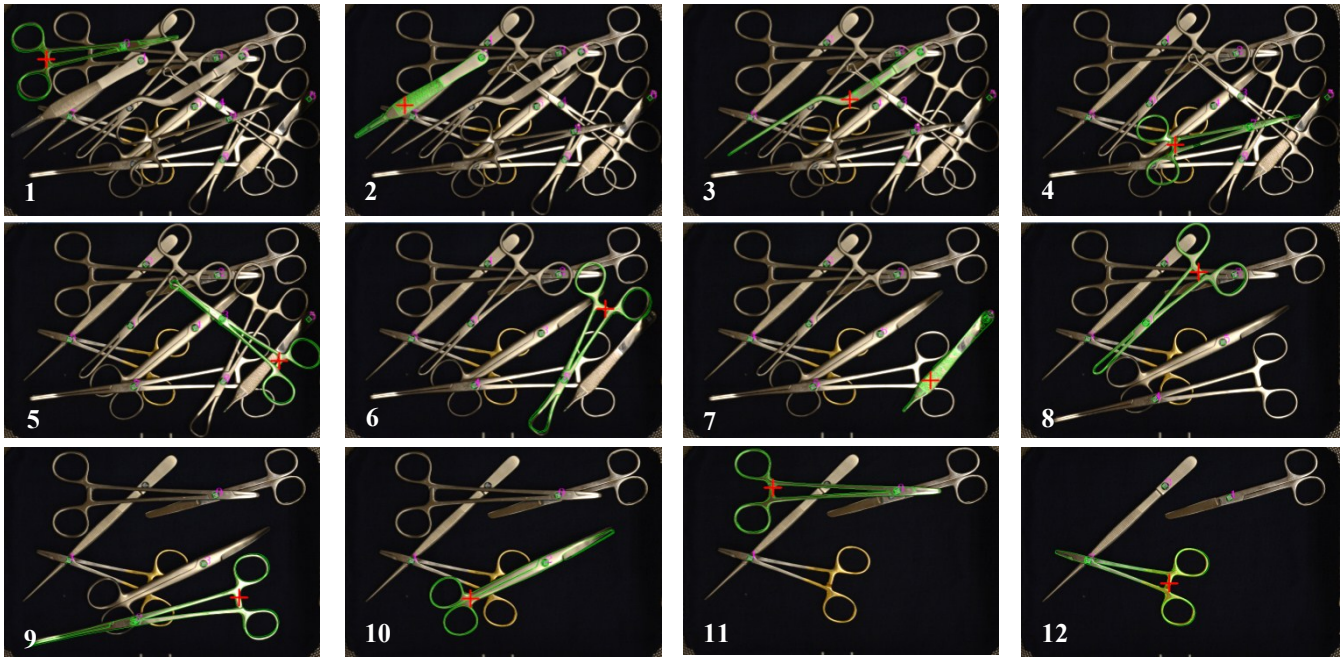


Figure 8. Camera views from an experimental run of our robot manipulator. Each image shows the container before gripping. The target instrument is highlighted in green. The 2D reference point for gripping is marked by a red "+". The final two images are omitted because there is no occlusion.

brings the instrument to a pre-defined pose and location for sorting purpose.

V. Experiments

We implemented our robot manipulator with an Adept robotic arm, a Nikon D5200 camera, and the custom designed electromagnetic gripper. The vision algorithm is performed on a Windows PC with 3.2GHz CPU and 8G RAM. To detect small 2D barcodes on the instruments, we use a DSLR camera with a 6000x4000 pixel resolution. In our experiments, we found that at least 70x70 pixels are required to decode a 2D barcode. Our occlusion reasoning algorithm is performed at a much lower resolution at 1500x1000 pixels. To suppress small edges and noise, we use a 7x7 Gaussian kernel to blur images before computing histograms.

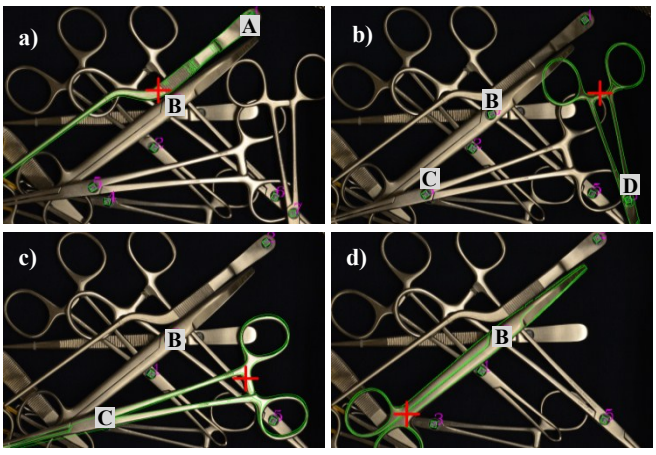


Figure 9. Recovery from a failed attempt. a) Failed attempt when trying to remove instrument *A*. b) Instrument *D* is determined to be occluded the least. c-d) *C* and *B* shifted and are removed subsequently.

We performed 15 experimental runs using our system. At the beginning of each run, the container is filled with approximately 12-15 surgical instruments. A total number of 192 images are captured and processed. The barcode detection on 6000x4000 images takes from 1.21 to 1.82 seconds depending on the number of barcodes in the view. The occlusion reasoning without pose optimization takes from 0.48 to 2.93 seconds to compute the gripping location depending on the complexity of the clutter.

Our vision algorithm successfully detects the topmost instrument or the one that is least occluded for 95% of the time (182 out of 192 images). When the candidate is correctly identified, the success rate for singulation is 98% using our electromagnetic gripper. Figure 8 shows the image sequence captured by the camera before each grip in one of the experimental runs. The next instrument to be removed is highlighted in green. The 2D reference gripping location is labeled using a red "+". The video attachment records the robot in action for this run.

Occasionally, the robot manipulator fails to grip a target instrument. This happens 3 times out of 182 tries. In Figure 9a, instruments *A* and *B* are aligned closely. The electromagnetic gripper makes contact with scissors *B* first and is not able to lift it up due to insufficient gripping force. The system decides to singulate the next available instrument. However, because instruments shifted during the failed attempt, there exists an occlusion cycle now in the scene (Figure 9b, *D* occludes *C*, *C* occludes *B*, and *B* occludes *D*). Since instrument *D* is least occluded, the system determines to singulate *D*. After removing *D*, instruments *C* and *B* shifted and are removed subsequently (Figure 9c-d).

When the vision algorithm predicts an incorrect instrument that is still occluded (10 out of 192 images), the

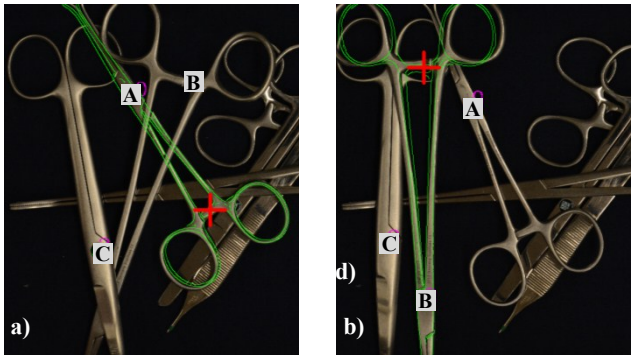


Figure 10. A case where barcode is occluded. a) Vision system determines instrument *A* to be on top incorrectly. b) After a failed attempt, instruments shifted; resulting in a new configuration where the vision system successfully finds the topmost instrument *B*.

gripper either grips a wrong instrument (2 times) or fails to grip due to insufficient force. For example in Figure 10a, instrument *A* is occluded by instrument *B*, whose barcode is occluded by instrument *C*. The vision system is not aware of *B*'s presence and determines *A* is on top of the pile. The system attempts to grip instrument *A* but fails, resulting in shifted instruments in the container (Figure 10b). *B*'s barcode is now revealed. The robot manipulator is able to recover from the previous failure and singulate instrument *B*.

VI. Conclusion and Discussions

We developed a flexible robotic manipulation system that is able to identify and singulate surgical instruments from a container. Our vision algorithm is robust against optical challenges such as changing light conditions, specularities, and interreflections among the surgical instruments. The design of a compliant electromagnetic gripper enables us to only solve a 2D pose estimation problem instead of more challenging 3D pose. Although the robot manipulator is designed for surgical instrument manipulation in this work, it can be extended to applications involving other mostly planar objects, such as certain industrial parts.

In the future, we will work on error handling. For the scenario in Figure 10, a vision-based object verification algorithm can determine whether the top object is occluded by some unknown objects. In the case where a wrong instrument is gripped, a verification step using an additional camera or by measuring weight can be incorporated to verify if the singulated instrument is the one determined by the vision algorithm. One potential disadvantage with using electromagnetic gripper is the tendency for the instruments to be magnetized over time. However, this can be addressed with the use of a standard demagnetizer (e.g., Neutrolator®). Currently, we require that all the instruments with a pivot (e.g., scissors, clamps) be in closed position. In the future, we would like to extend the algorithm to handle opened instruments. This involves identifying the instruments, finding the pivot location, and performing template matching on the two portions separately. We would also like to extend our system for picking up non-planar objects. We can achieve this by placing a barcode and creating a template at each stable position of the object.

Acknowledgment

This work is generously sponsored by the U.S. Department of Veterans Affairs Center for Innovation (VACI). Contract number is VA118-12-C-0051.

References

[1] RST PenelopeCS™, http://www.roboticsystech.com/
[2] K. Rahardja and A. Kosaka, "Vision-based bin-picking: recognition and localization of multiple complex objects using simple visual cues," in *IEEE/RSJ Int. Conf. Intell. Robots and Syst.,* Osaka, Japan, 1996, pp. 1448-1457.
[3] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3D geometry matching for grasping of known objects in cluttered scenes," *Int. J. Robotics Res.*, vol. 31, no. 4, pp. 538-553, Apr. 2012.
[4] M. Nieuwenhuisen, D. Droeschel, D. Holz, J. Stückler, A. Berner, J. Li, R. Klein, and S. Behnke, "Mobile bin picking with an anthropomorphic service robot," in *IEEE Int. Conf. Robotics and Automation*, Karlsruhe, Germany, 2013, pp. 2327 - 2334.
[5] C. Choi, Y. Taguchi, O. Tuzel, M.-Y. Liu, and S. Ramalingam, "Voting-based pose estimation for robotic assembly using a 3D sensor," in *IEEE Int. Conf. Robotics and Automation*, Saint Paul, MN, USA, 2012, pp. 1724 - 1731.
[6] S. Chitta, E. G. Jones, M. Ciocarlie, and K. Hsiao, "Mobile manipulation in unstructured environments: perception, planning, and execution," *IEEE Robot. Automat. Mag.*, vol.19, no.2, pp. 58-71, Jun. 2012.
[7] N. Shroff, Y. Taguchi, O. Tuzel, A. Veeraraghavan, S. Ramalingam, and H. Okuda, "Finding a needle in a specular haystack," in *IEEE Int. Conf. Robotics and Automation,* Shanghai, China, 2011, pp. 5963 - 5970.
[8] J. J. Rodrigues, J.-S. Kim, M. Furukawa, J. Xavier, P. Aguiar, and T. Kanade,, "6D pose estimation of textureless shiny objects using random ferns for bin-picking," in *IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, Vilamoura, Portugal, 2012, pp. 3334 - 3341.
[9] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T. K. Marks, and R. Chellappa, "Fast object localization and pose estimation in heavy clutter for robotic bin-picking", *Int. J. Robotic Res.*, vol. 31, no. 8, pp. 951-973, Jul. 2012.
[10] A. Pretto, S. Tonello, and E. Menegatti, "Flexible 3D localization of planar objects for industrial bin-picking with monocamera vision system," in *IEEE Int. Conf. Automation Sci. and Eng.*, Madison, WI, USA, 2013, pp. 168 - 175.
[11] A. Collet, M. Martinez, and S. Srinivasa, "The MOPED framework: object recognition and pose estimation for manipulation," *Int. J. Robotics Res.*, vol. 30, no. 10, pp. 1284 - 1306, Sep. 2011.
[12] X. Wang, T. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *IEEE Int. Conf. Comput. Vision*, Kyoto, Japan, 2009, pp. 32 - 39.
[13] S. Kwak, W. Nam, B. Han, and J. H. Han, "Learn occlusion with likelihoods for visual tracking," in *IEEE Int. Conf. Comput. Vision*, Barcelona, Spain, 2011, pp. 1551 - 1558.
[14] E. Hsiao and M. Hebert, "Occlusion reasoning for object detection under arbitrary viewpoint," in *IEEE Conf. Comput. Vision and Pattern Recognition,* Providence, RI, USA, 2012, pp. 3146 - 3153.
[15] A. M. Dollar and R. D. Howe, "The highly adaptive SDM hand: design and performance evaluation," *Intl. J. Robotics Res.*, vol. 29, no. 5, pp. 585-597, Feb. 2010.
[16] ISO, "Surgical instruments -- Metallic materials -- Part 1: Stainless steel," ISO 7153-1, 1991.
[17] Key Surgical KeyDot: http://www.keysurgical.com/
[18] Censitrac: http://www.censis.net/
[19] 2DTG Barcode Reading SDK: http://www.2dtg.com/
[20] B. Alefs, G. Eschemann, H. Ramoser, and C. Beleznai, "Road sign detection from edge orientation histograms," in *IEEE Intell. Vehicles Symp.,* Istanbul, Turkey, 2007, pp. 993-998.
[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. Comput. Vision and Pattern Recognition,* San Diego, CA, USA, 2005, pp. 886-893.