저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer
공학박사 학위논문

지능형 수술로봇 시스템을 위한 영상 기반 자동 상황 인지 기법에 관한 연구
(Vision-guided Automatic State-recognition Techniques for Intelligent Surgical Robot Systems)

2013 년 2 월

서울대학교 대학원
협동과정 비오엔지니어링 전공
류 지 원
지능형 수술로봇 시스템을 위한 영상 기반 자동 상황 인지 기법에 관한 연구
(Vision-guided Automatic State-recognition Techniques for Intelligent Surgical Robot Systems)

지도교수 김희찬

이 논문을 공학박사 학위논문으로 제출함
2013 년 1 월

서울대학교 대학원
협동과정 바이오엔지니어링 전공
류지원

류지원의 박사 학위논문을 인준함
2013 년 1 월

위원장

부위원장

위 원

위 원

위 원

위 원

서울대학교
Abstract

A number of robot-assisted surgical operation has been rapidly increasing over the past decade while the surgery with robots may have several potential issues. During the robot-assisted laparoscopic surgery, vascular injuries, which can be caused by mistreatment of surgical instruments, would be a major issue because the vascular injuries cause hemorrhage occurrences and tissue perforation that may be jeopardize the surgery itself. Limited vision through a laparoscope, limitations in agile re-configuration of the surgical instrument setup, or removal of bulky robotic tools and the surgical robot may hinder surgeons’ immediate emergency responses. To resolve this issue by providing preventive caution to surgeons, an advanced universal vision-based technique for object detection and tracking is proposed in this research. The suggested technique includes automatic detection of intraoperative hemorrhage and surgical instruments with depth information. This method locates the object region of interest by two common processes: feature extraction and tracking. Color and morphological information are used to segment the feature, and a Kalman filter is applied for robust tracking of the object locations with reduced error. Performance for hemorrhage and surgical instrument localization was quantitatively evaluated by root mean square error (RMSE) comparisons and instrument trajectory comparison with results.
of computerized methods and manual determination, respectively. Hemorrhage area variation analysis using proportionality of area increase and flow is also illustrated. Linearity of a positive slope for increased hemorrhage flow and negative slope for hemorrhage stanching were observed. Surgical instruments localization was evaluated through sensitivity and specificity measurement which were 86% and 96% respectively, and their depth information was validated through a simulation of their movement positions. It is concluded that the results are satisfying in the sense of over 80% state-recognition. Therefore, a vision-guided automatic state-recognition proposed in this study could minimize vascular injuries during the robot-assisted surgery.

**Keywords**: Hemorrhage detection, object tracking, robot-assisted surgery, surgical instrument tracking, 3D reconstruction, state-recognition

**Student Number**: 2010-31019
감사의 글

Acknowledgements

 언제나 제 곁에서 많은 도움과 격려를 아끼지 않았던 사랑하는 가족들에게 감사를 드립니다. 연구로부터의 성취감과 책임감을 모두 알 수 있도록 이끌어주신 김희찬 교수님. 존경하며 감사드립니다. 교수님의 지도하에 이 모든 과정을 마무리할 수 있어서 영광이었습니다.

 박사 논문의 심사를 맡아주셨던 김현회 교수님, 김성완 교수님, 최영민 교수님, 최재순 교수님께 감사드립니다. 교수님들께 느꼈던 카리스마와 포용력, 그 리더십의 강한함과 따뜻함을 항상 기억하겠습니다. 특히 박사 논문의 전 과정을 이끌어주셨던 최재순 교수님. 감사의 말씀을 말로 다 표현하지 못할 것 같습니다.

 연구 진행 방안에 대해 가르침을 주셨던 이정찬 교수님, 연구실 생활에 있어 여러 가지 도움을 주셨던 선배님들, 특히 석사 시절부터 큰 힘이 되어주셨던 지만선배, 만승오빠와 형선이에게 감사드립니다.

 아울러 박사 학위 논문에 조언을 아끼지 않았던 유진수군과 박사 과정동안 많은 도움을 주셨던 모든 분들에게도 깊은 감사의 말씀을 올립니다.

 2013년 1월
 류지원 올림
Contents

Figure Contents ................................................................. vi
Table Contents ................................................................. vi

1 Introduction ........................................................................ 1
   1.1 Robot-assisted Laparoscopic Surgery ................................. 1
      1.1.1. Background ............................................................. 1
      1.1.2. Importance and Needs ............................................. 7
   1.2 State-recognition in Robot-assisted Laparoscopic Surgery 10
      1.2.1. Importance and Needs ............................................. 10
   1.3 Research Objective and Specific Aims ................................. 17
   1.4 Organizations of the Dissertation ..................................... 18

2 Materials and Methods ...................................................... 19
   2.1 Universal Method Development ....................................... 19
      2.1.1. Classification .......................................................... 21
      2.1.2. Blob Analysis .......................................................... 28
      2.1.3. Object Matching ....................................................... 31
      2.1.4. Optimal Estimation ................................................... 37
   2.2 System Specifications .................................................... 46

3 Applications ....................................................................... 47
   3.1 Hemorrhage Recognition ................................................. 47
      3.1.1. Feature Extraction ................................................... 48
      3.1.2. Hemorrhage Tracking ............................................. 51
      3.1.3. Result ................................................................. 53
   3.2 2D Surgical Instrument Tracking ...................................... 61
      3.2.1. Feature Extraction ................................................... 61
      3.2.2. Instrument Tracking ............................................. 64
3.2.3. Result ................................................................. 69

3.3 3D Surgical Instrument Tracking ................................. 76
  3.3.1. Feature Extraction ............................................. 78
  3.3.2. Depth Extraction .............................................. 80
  3.3.3. Simulation ...................................................... 98

4 Discussion and Conclusion ........................................ 100

References ...................................................................... 104

국문초록 ..................................................................... 113
Table Contents

[Table 1] Strengths and limitations for surgeries done by humans and assisting robots ............................................. 9
[Table 2] Sensitivity and specificity of each surgical instrument detection ............................................................... 75

Figure Contents

[Fig 1] Da Vinci Surgical System ................................................................. 6
[Fig 2] Block diagram of the universal method ........................................... 21
[Fig 3] Color space converted surgical images ......................................... 25
[Fig 4] Kmeans clustered surgical images ............................................... 27
[Fig 5] Object blob analysis from surgical images ................................. 30
[Fig 6] Template matching for blood region and surgical instrument identification in surgical images .................. 36
[Fig 7] The concept of Kalman Filter ....................................................... 37
[Fig 8] FKF based object region tracking ............................................... 45
[Fig 9] Histogram equalization ............................................................... 49
[Fig 10] Binarization of hemorrhage occurrence image ....................... 51
[Fig 11] Results of hemorrhage detection ............................................... 54
[Fig 12] RMSE profiles of hemorrhage detection ................................. 56
[Fig 13] Hemorrhage flow classification by area variation analysis ........ 60
[Fig 14] Surgical instrument segmentation ............................................. 63
[Fig 15] Process of surgical instrument collision warning  ....  68
[Fig 16] Validation of surgical instrument position trajectories
in time domain .................................................. 70
[Fig 17] Validation of surgical instrument position trajectories
in spatial domain .................................................. 72
[Fig 18] Block diagram of 3D depth extraction and simulation
................................................................................. 77
[Fig 19] Image processing step of 3D surgical instrument
tracking ................................................................. 79
[Fig 20] Automatic Harris corner point detection .................. 82
[Fig 21] Feature matching via geometric threshold constraint 84
[Fig 22] Feature point normalization after Procrustes analysis
................................................................................. 88
[Fig 23] 3D reconstruction of surgical instrument via Delaunay
triangulation .......................................................... 94
[Fig 24] Depth extraction ................................................. 95
[Fig 25] Stereo vision block matching disparity map .............. 97
[Fig 26] Surgical instrument tracking simulation ................. 99
Chapter 1

Introduction

Vision-guided state recognition techniques for intelligent robot systems are considered in this dissertation. This chapter is devoted to the Introduction as follows: Section 1.1 describes the background of robot-assisted laparoscopic surgery. Section 1.2 addresses the importance and needs of state-recognition during robot-assisted laparoscopic surgery. Objects and the specific aims of the research is illustrated in Section 1.3 and the dissertation organization is explained in Section 1.4.

1.1 Robot-assisted Laparoscopic Surgery

1.1.1. Background

Robot-assisted minimal invasive procedures has been rapidly growing over the past decade. However, several potential issues remain in the surgery with robots.

Robot-assisted surgery mainly targets for minimal invasive surgery (MIS), a general surgical procedure avoiding large entry
incisions via long-handled instruments. The MIS is guided by visualization equipments such as an endoscope or a laparoscope. While an open surgery requiring a multi-centimeter incision, a laparoscopic surgery with three or four small holes, a single incision laparoscopic surgery (SILS) with one small incision, or a natural orifical surgery (NOS) with no external incision can be carried out [1-4]. Since the size of incision is quite small, the robot-assisted surgery provides the following benefits: increased accuracy, reduced assistants’ fatigue, less hospital stay time, etc. Technological benefits are further explained in Table 1.

Robot-assisted laparoscopic surgery (RALS) is a minimal invasive technique that miniaturizes surgical instruments and fits through a series of quarter-inch incisions instead of larger ones operated on patients [4].

The first robotic surgery was proceeded in April 11th 1985 as a CT-guided brain biopsy at the Memorial Medical Center, Long Beach, CA, USA using an industrial robot, Unimation PUMA 200 [6]. The idea of utilizing a mechanical structure held a guide in a correct position to lead the inserted probe to reach the surgical target without scratching vital tissues of the brain. This system was further developed to robotic devices, such as Vectorbot, an active robot by BrainLab (Feldkirchen, Germany) that had 7 degrees of freedom (Dof) where all electronics were
integrated into the articulated arm [7]. This robotic arm could be used to lock a jig or fixture in position so that a surgical tool could be inserted. In addition to these devices, TransUrethral Resection of the Prostate (TURP) and Probot with a purpose of autonomously removing a significant amount of tissues from a patient were developed in 1991 in London, UK [8].

The computer-assisted laparoscopic surgery field has been rapidly growing. Automated Endoscope System for Optional Positioning (AESOP), a voice-controlled to provide hands-free, ZEUS which is a master/slave robot with 4 doF, and a less dextrous version of da Vinci with a 6 doF were developed by Intuitive Surgical® (Sunnyvale, California, USA). Da Vinci system was first developed in the late 1980s to perform a remote surgery in the battle field. The Da Vinci surgical system is displayed in Fig. 1 where the technology was brought in 1995 to medical field, specifically in MIS field [5,9].

The Da Vinci system is programmed to reproduce the surgeon’s motions using a single or dual surgeon’s console in a master/slave configuration. The system configuration consists of three or four robotic arms, a stereo-visualization system of the surgical field, and proprietary instruments. The instruments mimic surgeons’ delicate hand motions of suturing, clamping,
and manipulating tissues, and filter tremors with scaled control from the console controls. Surgical assistants are by the side of the system for preparing the entry ports of the instruments and supervising laparoscopic arms. The instruments are inserted through 1-2cm ports of the patients, and the arms including an endoscope are localized according to the surgeon’s orders. The surgical view at the console is changed through the commands of zoom, movement, and rotation of the endoscope. The vision system uses high-resolution stereo-images that create a 3D view of the surgical field. It was sometimes claimed that a loss of haptic, a touch feedback, was crucial in this system, and efforts on researching developing haptic feedback were increasing [10-15]. However, until the technology is fully validated, the technology is replaced by more intuitive and improved visualization.

The commercial success of da Vinci is attracting new players. However, Intuitive Surgical’s success and dominance of the market acts as a barrier to entry as their installed base grows. As the number of procedures addressed by da Vinci increases, leaving less space for competitors. 25 years of age, according to Intuitive Surgical, 205,000 da Vinci-assisted procedures were performed in 2009, up to 51% from 2008 [5].

Several other recent technology that shows rapid advancement
in achieving minimal invasiveness includes Natural Orifice Translumenal Endoscopic Surgery (NOTES), which has advantages of leaving less scars since the instruments are inserted through the natural ports as the mouth or anus. Another rapid advancement in technology-assisted surgery includes Laparoendoscopic Single-Site Surgery (LESS) which the instruments are inserted through a single port providing a minimal scarring [16–17].

Further technology development research efforts are still in progress. Near future technology of surgical robots replaces repeatable work so that less assistants are necessary in OR, such as replacing a scrub nurse and a circulating nurse. This will create less dirt and bacteria-affected OR, so that it will create more safe environment. Moreover, technology will be able to tele-operate solders at war-site, and will may be headed to replacing fully automated robot to assistants, further to surgeons.

The aim for surgical robotics is to enhance the accuracy and shorten the procedure time, yet the improvement in quality produces high treatment costs. Such high costs increase the number of patients who can’t afford to pay, resulting in the main current issue. Seeking developments on the latest medical technology can potentially provide patients with the best
surgical outcome and clinical benefits.

Fig. 1.  Da Vinci Surgical System. (a) Insite view of da Vinci Surgical System [Intuitive Surgical®] (b) Surgeon’s view of the surgical site [Severance University Hospital, Seoul, Korea] (c) Surgeon console, patient cart, and robotic arms of the da Vinci Surgical System [intuitive surgical®]
1.1.2. Importance and Needs

The RALS has recently emerged as a viable MIS modality for the management of various urinary surgery [9–14]. About 98,000 cases of robot-assisted radical prostatectomy (RARP) were performed worldwide in 2010. The RALS has become the standard for prostectomy and radical hysterectomy in many centers and engaging popularity for colonoscopic perforation due to the need of dexterized surgical instrument handling and the advantage of minimal invasiveness [11, 19–21]. In addition, RARP provides decrease in surgical complication rates through less blood loss, improved transfusion rates, and fast return-to function [21]. Considering the anatomical structure of the internal iliac lymph nodes and the pelvic plexus, RALS with more extensive dexterity decreases the risk of damage to the small pelvis neural and vascular structures [22].

Comparing robot-assisted laparoscopic surgery to conventional open surgery, RALS has validated advantages in short hospital stay, less scar, and fast operation time [3]. Researches have shown that the longer time of operation causes the increase in surgeon’s mental and physical fatigue level. The result shows that during the surgery assistants will express less fatigue, leading to the better outcomes of the surgery. It also has shown that the result of robot-assisted surgery has enhanced lower
morbidity rates [23].

In technical point of view, the advantages of RALS over the conventional open surgery include the increased optimality in visualization of the surgical field, elimination of tremor, and increased motion and dexterity for suturing and manipulating tissues [11]. Table 1 displays the strengths and limitations for both surgeries operated by humans and robots.

RALS can also affect the surgical skills. Allowing the decrease in tremor and high scale of surgical movement, surgeon shows a great improvement in surgical skills. Superiority in surgical technique validation was achieved in previous researches through computerized assessment analysis [24-29].
Table 1. Strengths and limitations for surgeries done by humans and assisting robots.

<table>
<thead>
<tr>
<th>Humans</th>
<th>Robots</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strengths</strong></td>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td>- Strong hand–eye coordination</td>
<td>- Good geometric accuracy</td>
</tr>
<tr>
<td>- Dexterous (at human scale)</td>
<td>- Stable and untiring</td>
</tr>
<tr>
<td>- Flexible and adaptable</td>
<td>- Can be designed for a wide range of scales</td>
</tr>
<tr>
<td>- Can integrate extensive and diverse information</td>
<td>- May be sterilized</td>
</tr>
<tr>
<td>- Able to use qualitative information</td>
<td>- Resistant to radiation and infection</td>
</tr>
<tr>
<td>- Good judgment</td>
<td>- Can use diverse sensors (chemical, force, acoustic, etc.) in control</td>
</tr>
<tr>
<td>- Easy to instruct and debrief</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Limitations</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Limited dexterity outside natural scale</td>
<td>- Poor judgment</td>
</tr>
<tr>
<td>- Prone to tremor and fatigue</td>
<td>- Limited dexterity and hand–eye coordination</td>
</tr>
<tr>
<td>- Limited geometric accuracy</td>
<td>- Limited to relatively simple procedures</td>
</tr>
<tr>
<td>- Limited ability to use quantitative information</td>
<td>- Expensive</td>
</tr>
<tr>
<td>- Large operating room space requirement</td>
<td>- Technology in flux</td>
</tr>
<tr>
<td>- Limited sterility</td>
<td>- Difficult to construct and debug</td>
</tr>
<tr>
<td>- Susceptible to radiation and infection</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Table from Howe, RD, Matsuoka, Y. robotics for Surgery. Annual Review Biomedical Engineering. 1999, 01:213.*
1.2 State-recognition in Robot-assisted Laparoscopic Surgery

1.2.1. Importance and Needs

State-recognition is defined as an awareness of the surgical state, such as the recognition of hemorrhage flow, hemorrhage events, instrument collision, tissue perforation etc. Compared to the technical advantages of the conventional surgery, those of the surgical robots include the increased optimality, such as scaled visualization and resolution of the surgical field, elimination of tremor, increased motion and dexterity for suturing and tissue manipulation [11, 30-31]. However, as the complexity of the operating room increases, the medical professionals are forced to put more efforts into enhancing surgical safety. Thus, the rapid recognition of the surgical state during the operation reduces the chance of surgical complications [32-36]. If the injuries take place by the mechanical forces, the levels of medical complications become greater.

Robot system that dominates RALS market, da Vinci from Intuitive Surgical®, Inc., has limitations of a narrow surgical view through the scope, absence of force or tactile feedback, difficulties in agile tool management, and exchange of
instruments during emergency situations caused by the bulky and complex configuration of the system [29]. Because of the small workspaces and limitations in current robotic systems such as the absence of haptic feedback, robotic instruments may unintentionally exert excessive force on tissue possibly causing perforations, and instruments may collide and possibly compromise their mechanical structures during the operation.

Endoscopic vision–based danger state–recognition such as instrument collision detection, tissue perforation recognition, or hemorrhage detection, provides useful information for implementing safety mechanisms. In addition to the complications, computerized endoscopic video analysis can be beneficial for quantitative post-operative analysis of the surgery and the information archiving, and this method can also be utilized to automatically adjust a scope view [38-40].

In order to help surgeons work with extra preventive caution during surgery, a vision-guided technique for automatic object detection in endoscopic video is introduced to facilitate faster awareness of the surgical state.

Object detection during the surgery is significant based on the fact that the surgical system is based on the mechanical product. With the aid of detecting instruments and other
situations during the surgery, the late response to accidental injuries during the surgery that may cause problem in immediate responsive reconfiguration of the surgery robot setting will be prohibited [29, 40-42]. However, only few studies on the warning system based on the state-recognition have been actively explored. To increase the reliability of the robotized laparoscopic surgery, computer vision techniques can be applied to the automatic recognition of the current surgical states, such as tissue perforation and surgical instrument localization.

**Hemorrhage Recognition**

Hemorrhage is one of common complications that may randomly arise during surgery. An immediate response to vascular injury and unexpected hemorrhage that may result in fatal process of hypovolemia is extremely significant [43].

According to Blum et al. [44], uncontrollable bleeding from the cystic artery in cholecystectomy results conversion from laparoscopic to open cholecystectomy. Also, issues on post-operative hemorrhage events from vascular complications are remained as huge problems in robot-assisted surgery [45]. A review of urological reports shows that a vascular injury rate
laparoscopic surgery is 0.03~2.7% and a recent review reported that an intraoperative complication rate of less than 5% in laparoscopic surgery. However, hemorrhage during and after the surgery accounts for 40% of all perioperative complications [46-49]. These kinds of hemorrhage, mainly occurred after the contact of robotic equipment, flow out gradually leaving holes to the tissues, so that the surgeons can easily miss out hemorrhage events and end up facing a difficult post operation. Such a failure in a surgery can be fatal when the particular tissues should be injured, such as duodenum or iliac artery in prostate surgery. Even though vessel injury occurrence is lower than that of other complications, the mortality is relatively high. Injury to one or more major vessels can quickly result in fatal exsanguination, leading to most of deaths occurring within the first 24 hours [50]. Delayed post-operation hemorrhage directly leads to medical complications and even result in the loss of an organ [41-42]. Thus, hemorrhage detection is an extremely serious issue.

Because of the limited view through a laparoscope and limitations in the emergency response of a robotic system regarding the detection of hemorrhage, surgeons confront difficulties in detecting hemorrhages and controlling incidental injuries during surgery.
In technical point of view, most of the past researches deals with the application of vision techniques focused on the segmentation of the blood region in capsule endoscopy, detection of hemorrhage in fundus images, image-based instrument tracking during robot-assisted surgeries for automatic laparoscope view adjustment, etc. [51-52]. Automatic detection of bleeding in capsule endoscopy images was implemented with a color texture feature, and hemorrhage in digitized fundus images was detected based on brightness correction.

Even though hemorrhage detection during robot-assisted surgery has a worldwide interest because a mechanical force of the robotic arms damaging the soft tissues, not many researches about the hemorrhage detection have been performed. During the surgery, events seen through the surgical view can be mapped and recorded. This can further be implemented to act as a tissue perforation prohibition warning that may warns when foreign body enters the delicate tissues that is fatal when hemorrhage occurs. In previous studies, the methods dealt with images that were not on-line, where the region-of-interests (ROI) do not need to be traced in real time. Therefore, the suggested techniques did not apply to intraoperative hemorrhage during surgery.
Surgical Instrument Tracking

In addition, virtual wall or fixture implementation based on instrument tracking form a virtual repulsive force field around the instrument, and can be an example of the safety mechanism providing surgeons with instrument collision warning for safer operation. This type of the safety function is more meaningful in surgical tasks in which skills with lower dexterity increase the risk of tissue perforation, for example, during right lower lobectomy or other surgeries, main veins and arteries should be cautiously managed to prevent further injuries [53]. Stradler P. et al. [54] has claimed that among 150 patients who have received vascular surgeries, 4 patients, 2.7%, experienced serious post-operative complications from the surgical equipment. Studies have shown that tissues can be overloaded from the surgical instruments, which can lead to a serious complications [55]. As well, vessel injury occurrence usually due to surgical instruments damaging one or more major vessels results a fatal tissue perforation with high mortality rate [50].

In order to warn the surgeons when specified tissues are about to be damaged by surgical instruments, the accurate surgical instrument localization is needed to be proceeded. Most current surgical guidance systems use additional markers to trace
surgical instrument position, whereas surgeons acquire most information from visual feedback. Localizers based on optical, electromagnetic, mechanical, or sonic technologies track instrument markers outside the body, and the instrument positions are estimated according to the markers attached to the instrument [66–67]. However, because of the surgeon’s movements and occlusion by the camera or detector, the position of these localizers in a cluttered operating room becomes a critical issue. As an alternative, vision-based approaches using artificial color markers attached to the instruments as in vivo sensors and vision processing algorithms have been attempted [38–39, 59–65]. However, these efforts have increased concerns about biocompatibility and sterilizability, and setting the markers on all instruments can be troublesome and may slow down the procedure during switching between instruments.
1.3 Research Objective and Specific Aims

In technical point of view, a novel method that does not require the additional hardware but depends only on visual feedback from the actual surgical environment is needed to be developed to overcome the drawbacks of conventional approaches of surgical robots. Enhancement of safety during the surgery via novel functions such as hemorrhage detection caused by tissue perforation due to the contact of instruments, and an instrument collision warning based on robust real-time tracking of multiple instruments can be achieved. Furthermore, stereovision based 3D reconstruction method is considered as a good option to further provide depth and volumetric information of surgical instruments.

Thus, our research goal is to the development of a new functionality of vision-based automatic state-recognition utilizing a common algorithm framework that can be applied to various objects in the endoscopic video during robot-assisted surgery. We propose a universality with an on-line and real-time state-recognition algorithms. Applications including hemorrhage detection and multiple surgical instrument tracking from the laparoscopic images obtained during robot-assisted surgery was achieved considering the importance of the needs.
1.4 Organizations of the Dissertation

Chapter 1 handles an introduction of the dissertation. The background of the surgical robot operations and state-recognition technologies are introduced. The methodology of the state-recognition technologies to overcome the potential issues during surgical site is illustrated in Chapter 2. The method includes an object segmentation and tracking algorithms which are described with two types of surgical objects: hemorrhage and surgical instruments. Chapter 3 deals with the object recognition in surgical view. Specifically, applications regarding hemorrhage and surgical instrument detection were implemented. Chapter 4 discusses and concludes the dissertation.

In this chapter, the background and needs have been described. The objects and the specific aims of the dissertation and its organizations have been also addressed. A development of state-recognition method will be followed in Chapter 2.
Chapter 2

Materials and Methods

2.1 Universal Method Development

To cope with difficulties in the processing of laparoscopic images that contain high speculation noise and non-uniform background, the proposed universal object tracking algorithm was structured with two characteristic stages: 1) feature extraction and 2) object tracking.

Universal object recognition method can be defined as a common algorithm that can be applied for various object detection providing satisfactory results. Despite the fact that automatic object detection in endoscopic video or other medical vision has been researched, and drawing widespread interest in various applications, little literature is found regarding common framework or universal algorithm that can be used to various object detections concurrently. The object detection algorithm available now subordinates to a specific system, detecting specific types of object. Moreover, the object detection algorithm in various field is usually developed for only for research
purposes, thus limiting the implementation in real surgeries. The level of object detection technology during robot-assisted surgery now is at the early stage. Due to the increase use of robot-assisted surgery, and image-guided surgical robots from the various providers, standardization and fast adaption of the algorithm during real robot-assisted surgery is needed.

The universal method used to identify the hemorrhage region and surgical instruments in two-dimensional (2D) images is based on local image features and accounts for the color and morphological properties of the objects. Fig. 2 shows the block diagram of the proposed detection method which composed of three main processing phases: 1) segmentation using specific color properties of each image frame, 2) template matching to localize the object using the correlation between image frames, and 3) optimal estimation of the current object locations using the Kalman filter [68].

The proposed method can be propagated as the universal method since the method contains all kinds of image processing analysis, including spatial and temporal information, color and intensity information as well as optimal estimation. Hence, the proposed method can work as a base framework for an object recognition system in surgical images.
2.1.1. Classification

Color Space Conversion

A conversion of the best-fit optimal color space is required to classify the region of the interest accurately. The processing of surgical images uses three of the color spaces which are RGB (Red-Green-Blue), LAB (Lightness and color-opponent dimensions), and HSV - (Hue-Saturation-Value) spaces.
The provided surgical images are in RGB space as general digital color images. RGB consists of three dimensions of red, green and blue. Such space provides better segmentation results than those of any other color spaces due to sharp color definitions [66]; however, changes of each dimension values are not perceptual as human eyes.

In order to provide perceptual imaging, image color conversion to HSV and LAB spaces is provided, respectively. Hue, H from HSV, stands for the color type which has a range from 0 to 255 with 0 being red. Saturation, S, represents the vibrancy of the color, ranging from 0 to 255. The lower the saturation value, the more gray color presents, causing an image to appear faded. Value, V, represents the brightness of the color, ranging from 0 to 255 with 0 being complete darkness and 255 being full brightness. HSV model is a common cylindrical-coordinate representation of points in an RGB color model a simple transformation from device-dependent RGB space. The conversion of RGB to each HSV space is shown in (1) to (4), respectively. Therefore, HSV color space is known to be more intuitive, and the changes toward the light source are not as dramatic as those in RGB color space [66].
\[
H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B \geq G \end{cases}
\]  \quad (1)

where

\[
\theta = \cos^{-1} \frac{1}{2} \frac{[(R - G) + (R - B)]}{[\sqrt{(R - G)^2 + (R - B)(G - B)}]} \quad (2)
\]

\[
S = 1 - \frac{3}{(R + G + B)} \min(R, G, B) \quad (3)
\]

\[
V = \frac{1}{3} (R + G + B) \quad (4)
\]

However, HSV color space is not independent from the light source. To overcome its lack of perceptual uniformity, more computational intensive model LAB space is introduced.

Unlike RGB color space, LAB color space is designed to approximate human vision so that lightness chromacity, L, closely corresponds to human perception of lightness. L in LAB color space stands for lightness, values from 0 to 100 each displaying black to white, A dimension is a band between the green space and magenta, and B dimension represents the space between yellow and blue. Because LAB space has larger gamut than previously mentioned color spaces, 16bit/channel computation is needed instead of 8bit/channel to overcome the
data loss during the operation.

This color space conversion, displayed in Fig. 3, is useful in pre-operation process of object detection in which the object of interest could be distinguished from environments using the contrasted color signatures. Utilizing the advantages of each color spaces allowed k-means clustering as one of the post-algorithms using LAB spaces which convey the levels of red chromacity and luminance to classify the object.
Fig. 3. Color space converted surgical images. (a) red, green, blue image from RGB space starting from the left (b) Hue–Saturation–Value (HSV) image starting from the left (c) LAB image starting from the left.
**K-means Clustering**

After color space conversion from device-dependent RGB space to LAB space, objects are classified into two groups using k-means clustering. In each dimension of LAB space, the definite separation occurs in L and A spaces where L indicates the luminescence of the image and A distinguishes organs from surgical instruments using red chromacity.

K-means clustering is a distance-based clustering model that groups data into a given number of dataset by calculating the sum of squares between the temporary centroid and input data.

The algorithm is a popular algorithm for unsupervised learning of Neural network [67]. This can be implemented in high dimensionality data such as surgical images in order to classify background with foreground, malignant tissues with normal tissues, and, in our case, surgical instruments with other organs using intensity values of each dimension. This is represented in Fig. 4 where the surgical instruments were differentiated from the organs.

There are other similar algorithms such as Expectation Minimization (EM) algorithm that calculates the maximum likelihood of the data sets to identify groups in an image.
However, EM fails on high-dimensional data sets due to numerical precision problems [69]. Thus, k-means provides better performance than other classifiers in high dimensional data as the surgical videos.

Fig. 4. Kmeans clustered surgical images (a)(b) classification of surgical instruments (left) and organs (right) via kmeans clustering
2.1.2. Blob Analysis

Blob detection segments objects that differ in properties such as brightness or color from the surroundings and includes corner detection and interest point detections. The blob analysis is usually used for further processing to find ROI and object recognition from analyzing texture. Blob analysis in this context contains the chain of image processing techniques and image filters.

Applying the blob analysis, color converted image or the object of interest classified from k-means clustering algorithm goes through a chain of image processing techniques for noise removal. The techniques involves erosion, dilation, hole-filling, small island removal, threshold technique, canny edge detection, entropy filtering, etc.

Threshold technique defines color differences of the object of interest from the environment. Following the threshold technique which creates a gray scale or a binary image, entropy filtering remove the speckle noise from the background utilizing the relative changes of entropy, where the output is a blurred grouped object image. In (5), \( H(L) \) denotes the output entropy, and the entropy of the belief over \( L \) is defined as \( \text{Bel}(L) \) [70].
Entropy filtering implemented in our image redefines the image by replacing every pixel entropy value in the image with the mean entropy of the values in 9x9 neighborhood. This entropy filtering creates an entropy filtered obscure image with speckle noise removed.

Canny edge detection then defines the edge of the closed contour so that the object boundary and the number of objects can be extracted. By using binary edge detection, we are able to eliminate the image details completely and maintain the shape features.

To represent how previously mentioned algorithms affect the surgical images, the outcomes of entropy filtered image and canny edge detected images process is displayed Fig. 5 where RALS image extracted from da Vinci system was used.
Fig. 5. Object blob analysis from surgical images. Left column denotes surgical instrument blob analysis, and right column denotes blood region identification. (a) original image (b) canny filtered image (c)entropy filtered image with input from (b) (d) noise eliminated output image from the blob analysis
2.1.3. Object Matching

Template Matching

Because the segmentation process sometimes fails owing to occlusions by other objects and noise under time-varying light conditions due to the properties of surgical images, a similarity measure algorithm, such as template matching, also needs to be applied simultaneously.

Template matching can be classified into two categories: feature-based and template-based. Feature-based approach utilizes edges and corners of the template to find the best-matching location in the source image. On the other hand, template-based approach used the entire template image and with a sum-comparing metric such as sum of absolute difference (SAD), the sum of squared difference (SSD) or the normalized cross correlation (NCC) within search image to determine the best location match of the template.

In this paper, template-based approach, so called global approach, is applied to find the best match of an object-of-interest in timely manner. To recognize the object movement of shape-change, template is given as the defined segmented object, and the source image is set as the
after-frame of the frame that contained the template image. To find the best match, minimizing SSD measurement was chosen. This process is represented in (6) where $f$ is the candidate image, and $t$ is the template image. In our method, $f$ was set as the next frame surgical image, and initial $t$ was the object region segmented from the first step of the algorithm. $t$ is kept updated as the result from the template matching. Summation was computed over positions $x, y$ under the template positioned at $u, v$.

$$E_d(u,v) = \sum_{x=1}^{N} (f(x,y) - t(x-u, y-v))^2$$  \hspace{1cm} (6)$$

By calculating SSD for each pixel of the two images, the highest correlated position, which is the lowest value, can be obtained and set as the center of the match block.

Showing more complex calculation when using normalized cross correlation (NCC), NCC has the advantage of independence to the illumination but fully depends on the texture of the image. To simply explain, NCC matches images via the edge lines of the objects in images. However, due to speculation noise and surgical environment characteristics as in non-smooth background, NCC does not fit for our purpose.
Since pre-processed image data filters illumination, simpler calculation with the accurate correlation measurement works better for the surgical images. The comparisons of SSD and NCC are shown in Fig 6.

**Fast Fourier Transform based Template Matching**

Before implementing template matching, converting image using fast Fourier transform (FFT) results images in frequency domain. Since template-based template matching may potentially require large number of sampling points, using FFT is efficient for the template matching algorithm to overcome the slow computation speed rather than applying template matching in the spatial domain [71].

The Fourier transformed image decomposes an image into sinusoidal components. It is easier to examine the certain frequencies of an image, thus influencing the geometric structure in the spatial domain. Also, applying FFT on an image results a greater range than the spatial imaged domain, thus the values are usually stored as float numbers. FFT is used to compute Fourier transform with less number of calculations. It creates the same results but reduces the computation time by hundreds, and the complexity reduces from $N^2$ to $N\log_2 N$ due to the calculation similar to the method called “divide and conquer.”
The number of frequencies in FFT image corresponds to the number of pixels in the spatial image, so the sizes of the image in spatial domain and frequency domain are the same. Given N x N template image sized two dimensional Fourier transform, where \( f_T(x,y) \) is the spatial domain template image, \( F_T(k,l) \) corresponds to the Fourier transformed template image which is the sum of the calculation of Fourier transform on each pixel of the spatial domain template image as represented in (7). The same fourier transform process for the reference image is implemented using (8). MxN reference image \( f_R(x,y) \) is fourier transformed into frequency domain image \( F_R(k,l) \).

\[
F_T(k,l) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_T(x,y) e^{-j2\pi \left( \frac{k x}{N} + \frac{l y}{N} \right)} \tag{7}
\]

\[
F_R(k,l) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f_R(x,y) e^{-j2\pi \left( \frac{k x}{M} + \frac{l y}{N} \right)} \tag{8}
\]

Since the results are interpreted as sine and cosine waves, \( F(0,0) \), the center of the image, showing the average brightness and \( F(N-1, N-1) \), which is the furthest location from the image center, represents the highest frequency. As Fig. 6(b) demonstrates, the image contains the components of all frequencies, but their magnitudes get smaller for the higher frequencies. The low frequencies contain more image information than the higher ones do. The two dominating directions, vertical
and horizontal line in the figure, denote the regular patterns in the background of the source image.

Applying FFT to both template and the source image, SSD is calculated using both magnitude and phase of the FFT images. To make the computation faster, downsizing the sample window or reducing image resolution can be the options. Thus, downsizing the source image to the similar locations of the template image, the computation time decreases since the template image represent the segment image in the previous frame.

To convert the frequency domain image back to the spatial domain in order to locate the spatial location of the matching point, the inverse of Fourier transform is applied using (9), where $f(a,b)$ denotes image in spatial domain, and $\frac{1}{N^2}$ represents normalization.

$$f_R(a,b) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F_R(k,l) e^{j2\pi \frac{ka}{N} + \frac{lb}{N}}$$  \hspace{1cm} (9)$$

Implementation of the template matching algorithm overcomes the drawback of local color-dependent segmentation which may not work under different lighting conditions.
Fig. 5 Template matching for blood region and surgical instrument identification in surgical images. Each row indicates each object matching: first row: blood region identification second and third row: surgical instrument identification (a) blood region and surgical instrument template used for template matching (b) sum of squared difference matching image after FFT transformation (c) normalized cross correlation matching image after FFT transformation (d) result of blood region and surgical instrument detection, positions of two instrument combined in an image, from the template matching process.
2.1.4. Optimal Estimation

In this study, a Kalman Filter [68] is being applied to get the optimal estimates. Since the Kalman Filter is very important, a set of Kalman filter is revisited in this section.

Kalman Filter

Fig. 7. The concept of the Kalman Filter [68]

After the appearance of the application of segmentation process and template matching algorithm, Kalman filter (KF) technique was utilized to compensate for tracking failures such as a detection failure due to occlusion of smoke or foreign body. When both previously-mentioned stages fail to correctly detect the object-of-interest, the KF functions as an estimator of the object position. The object movement is hypothesized as a linear
model, and thus linear KF tracks the object. The input of the filter was set to the closest point as the previous Kalman filtered position, and hence, KF built on inputs of acquired knowledge from previous object locations will provide more accurate and robust results.

KF is a linear, recursive estimator that produces a statistically optimal estimate from calculating a minimal mean square error (MMSE) of the underlying system state. KF uses a series of measurements observed over time, under the assumption of data containing white, Gaussian noise. In other words, KF is used to estimate results by eliminating uncertain noises from previously collected data and newly obtained data. Because of the property of the unbiased estimate, KF tends to be more precise than those based on a single measurement alone; also, the recursive property provides a real time application of KF. From a theoretical standpoint, the main assumption of KF is that the object movement in nature can be expected, and a property of movement is natural.

Many filters depend on the frequency level, such as low pass filter, high pass filter, or band pass filter. However, obtaining the optimal results from non-physical filtering in frequency domain using digital signal processing calculations is difficult to achieve. Hence, processing in time domain is necessary. Signals
in image have noises from image sensors, such as smear, constant noise, and low light level noise, which are included in sensor signals. KF includes two groups of noises: a noise from the image sensor which are usually constant, and the one from the measurement, so called a bias. We hypothesized that quantization error, which describes the error when the continuous signals are converted to pixel-wise format, and movement position detection error are included in the measurement.

KF in this paper is a linear system. In other words, the system model can be represented with linear operators. Linear systems are difficult to find in reality so that most of non-linear systems use the Extended Kalman Filter or Unscented Kalman Filter. However, we hypothesized that the object movements, such as surgical instrument movement and hemorrhage flow, has its physical inertia as it could be modeled as linear system. Therefore, the system can be approximated and written as in $y = ax + b$ form.

The process model which represents the system state propagation and the measurement model measured directly from the sensors are shown in (10) and (11) respectively. The process model is usually designed by the user according to the system’s physical characteristics. In (11), $z$ denotes the sensor
measurement vector, \( \mu \) defines the measurement noise, and \( s \) from the process model confirms the \( z \) value. \( Bu \) is a user control input model for controlling the action, and in the current value \( s \), the process noise \( w \) which contains the noise that the system cannot control such as the white noise is summated in (10). \( A, B, H \) are a process model matrix, a control input model matrix, a measurement model matrix, respectively, and \( k \) denotes the time index. The desirable result is the estimate of the value of \( s \) in time \( k+1 \).

\[
s_{k+1} = As_k + Bu_k + w_k \quad (10)
\]
\[
z_k = Hs_k + \mu_k \quad (11)
\]

In the implemented algorithm, 4 states - \( x, y \) position and velocity estimation outputs from the KF, and the input sensor measurements were given as 2 for the segmentation and template matching measurements. When the template window size in the template matching step is over 100x100 pixels, an offset between the outputs from the segmentation and template matching occurs. Then, the states would include the offset positions and become 6 states, and the measurements would be given as 4. The initial estimate setting for \( \hat{s}_{k-1} \) and \( P_{k-1} \), which denotes the previous state estimation error covariance matrix in (12) and (13), were given as 0s as an input, and an identity
matrix of ∞ denoting uncertain initial value, respectively. The variance of the estimation error variance is defined by (14). This initial setting was applied for a rapid convergence of the state estimation. With given the estimate of the state ahead and process error covariance, the current state was predicted using (12). Bu was set to 0 because no additional input was known for our system which was for the detection of an object rather than the control of the mechanical actuator.

\[
\hat{s}_k = A \hat{s}_{k-1} \\
P_k^{-} = A P_{k-1} A^T + Q \\
P_k = \sigma_k^2 = E[(z_k - s_k)^2]
\]  

In (12) through (18), \( \hat{s} \) is the \textit{a posteriori} state estimate, which denotes the updated estimate and \( \hat{s}^{-} \) represents the \textit{a priori} state estimate which is the predicted state estimate. \( Q \) in (13) is the process noise covariance, which was set as 4x4 identity matrix, and \( R \) is the measurement noise covariance which was set as 2x2 identity matrix modified according to the measurement input value. The estimated kalman filter positions and each measurements were compared, and the \( R \) were tuned depending on the distance error.
In (15), $\gamma_x$ and $\gamma_y$ denotes the tuned measurement error of localization. Then, the kalman gain, $K$, was kept updated to adjust the weight of the kalman filter estimate.

\[
R = \begin{bmatrix}
\gamma_x & 0 \\
0 & \gamma_y
\end{bmatrix}
\]

In (15), $\gamma_x$ and $\gamma_y$ denotes the tuned measurement error of localization. Then, the kalman gain, $K$, was kept updated to adjust the weight of the kalman filter estimate.

\[
K_k = P_k^{-1}H^T(HP_k^{-1}H^T + R)^{-1}
\]

\[
s_k = \hat{s}_k + K_k(z_k - H\hat{s}_k)
\]

\[
P_k = (I - K_kH)P_k^{-1}
\]

In (16), the measurement update to minimize the error using the estimate and the real measurement value, $K$ was computed, and the estimate, (17), was updated. $\hat{s}_k$ with the measurement $z_k$, and the error covariance, $P_k$ was finally updated for optimality. (12) to (18) worked as a closed loop and iterated throughout each time frame. The overall block diagram of the KF concept is depicted in Fig. 7.

The basic hypothesis of KF is that the measurement noise is time-independent white noise and can be represented as Gaussian noise. Based on this assumption, weight gain on the object color segmentation and template matching depends on the distance measure of the previous position. Application of KF
achieves smoother tracking position localization and refines the occlusion problems.

**Federated Kalman Filter**

The Federated Kalman Filter (FKF) is usually implemented in sensor fusion field. FKF consists of a combination of several local filters and one master filter, where this is an advancement of the general KF [72]. As shown in Fig. 8, the first part of the FKF has two local filters applied to the segmentation, and template matching. Each independent measurements are processed by each local filter reducing any processing error during each operation. Although the measurements are independent of each other, the KF estimates from different measurements are correlated due to the common noise states of the inertial system model, (12) [73]. The outputs from the local filters are applied to a master filter to finalize the location of the object position.

The same Kalman filter parameters for initial measurement and process error parameters are used for the FKF. With the modification of each measurement error update depending on the input object, the final state estimate is calculated.

Generally, KF is known to provide high accuracy without large
amount of calculations, which is not popular for real-time object tracking [74]. To overcome the limitations of KF in non-linearity and non-gaussian noise, we have provided adjustable weight gain and noise model using Euclidean distance measure of object position in time domain as well as between each method.
Fig. 8. FKF based object region tracking.
2.2 System Specification

Various types of recognitions required the specific numbers of videos: Three for hemorrhage state-recognition, four for prohibited region warning through multiple surgical instrument tracking, one stereo video for 3 dimensional (3D) surgical instrument reconstruction for depth information. Robot-assisted laparoscopic surgical operation videos had specifications of 640x480 pixels in HD resolution and approximately 20 to 30 frames/s in MPEG and AVI format. Each video consisted of a sequence of different surgical tasks under numerous lighting conditions. The processing and analysis steps were implemented using MATLAB (MathWorks, Natick, MA, USA). Each video for hemorrhage state-recognition covered a time period of approximately 20s with two of the videos containing hemorrhage occurrences and 1 min video for surgical instrument tracking.

Computer specifications to run the program were done with i7 CPU, NVIDIA Quadro FX1800 graphic card, and 4G memory.
Chapter 3

Application

3.1 Hemorrhage Recognition

In this paper, we provided a safety-assured functionality of hemorrhage state-recognition to surgical robot systems through image analysis. Developing a robust segmentation and tracking of hemorrhage recognition in color laparoscopic image using a machine vision system was a challenging tasks. The laparoscopic images are usually under time-varying lighting conditions, scene change, and a moving background due to pulsation and breathing. The hemorrhage region may be occluded from the surgical view by the surrounding organs or surgical instrument. Thus, detection mechanism should be performed as well as tracking algorithm in order to provide surgeons a correct location of hemorrhage. Color signatures, image factors, and motion filtering were used to provide a reliable, fast, and robust detection and tracking algorithms:

The proposed method to identify hemorrhage region in two-dimensional surgical images was based on the local image
features such as color and morphological properties of hemorrhage. The three main processing phases were introduced in the block diagram of the universal method in Fig. 2: 1) hemorrhage segmentation with specific color properties of each image frame, 2) template matching to locate the hemorrhage site employing the correlation between image frames, and 3) optimal estimation of the current hemorrhage location using Kalman filtering.

3.1.1. Feature Extraction

The first stage of the application of the universal method was to use different color signatures of hemorrhage from surrounding backgrounds such as organs and instruments. Under reasonable surgical endoscopic lighting conditions, hemorrhagic blood possesses a distinct color signature of blackish red from other body tissues and instruments. In order to correctly distinguish a hemorrhage region, the image contrast was enhanced by using histogram equalization in RGB space.

Histogram was used to adjust image contrast. This algorithm was effective when data in images have close contrast values. Through the equalization adjustment, images were normalized and mapped from the local highest value to 255. If $k^{th}$ pixel
from $T$ input image goes through histogram equalization, due to normalization and the fact that all pixels have to receive a value, the equation would be (19).

$$I(k) = 255 \times \frac{\sum_{i=0}^{k} \text{histogram}}{\sum_{i=0}^{n} T}$$

(19)

where $I$ is the output of histogram equalization, and $n$ is the number of pixels. Histogram equalization for each R, G, B, space was calculated and combined.

**Fig. 9. Histogram equalization.** (a) original image of hemorrhage occurrence (b) histogram equalization applied hemorrhage occurrence image

Using the contrasted image shown in Fig. 9, a mutually inclusive threshold technique expressed in (20) was implemented
to identify the hemorrhage region. This algorithm was applied to optimize a threshold to each RGB space and extracted hemorrhage regions by combining the mutually inclusive points, producing a binary image $I$.

$$I = R_{\text{threshold}} \cap G_{\text{threshold}} \cap B_{\text{threshold}}$$  \hspace{1cm} (20)

where $\cap = "\text{and}"$

The canny edge detection technique and entropy filtering were applied thereafter to find a closed contour and to decrease the background noise. The image details and the shape features could be completely eliminated and maintained respectively using binary edge detection, as shown in Fig. 10. Finally, the segmented boundary of the hemorrhage region was marked on the original image.

Due to the characteristics of the hemorrhage where blood of vein and artery posses different color, this feature extraction algorithm may work as a universal algorithm for various hemorrhage detections. This algorithm is not just restricted to hemorrhage from a specific vessel; moreover, different kinds of hemorrhage can also be detected through modification of the hemorrhage detection characteristics.
3.1.2. Hemorrhage Tracking

Because the segmentation process sometimes fails due to occlusions by other objects and noise under time-varying light conditions, a template matching algorithm was also applied simultaneously as shown in Fig. 6. The average location of previously identified regions that were sampled at approximately 0.5s was used as a template, in which the window was sized approximately 5.5 % of the reference image size. This method was groundbreaking because of its ability to track the hemorrhage region that could not be distinguished by background colors of local features.
The highest correlated position expected to be the hemorrhage position in time domain could be obtained by calculating the minimum SSD value. By reducing the candidate image region from the entire picture to the position around the hemorrhage area and using a FFT based calculation algorithm, the slow computation speed could be overcame.

At the final stage of the algorithm, a FKF was applied to the result of the segmentation and template matching processes to optimally locate the hemorrhage region. The first part of the FKF had two local filters that were applied to the segmentation and template matching results of hemorrhage detection independently to reduce any processing error during each operation. Using the same KF model represented in (21), \( \Delta T \) denoting the sampling time, and the measurement model in (22), the outputs were calculated. Then, the outputs from the local filters were applied to the final master filter weighted measurement noise update. By calculating Euclidean distance between the two local filter outputs from the previous hemorrhage location, the input position was set, and the measurement noise matrix was updated. The positions were used to finally locate the hemorrhage region. The designed FKF model for hemorrhage region tracking is displayed in Fig. 8.
\[ s_{k+1} = A s_k + w_k \]  \hspace{1cm} (21)

\[
\begin{bmatrix}
1 & 0 & \Delta T & 0 \\
0 & 1 & 0 & \Delta T \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x_k \\ y_k \\ x_k' \\ y_k'
\end{bmatrix} + w_k
\]

\[ z_k = H s_k + \mu_k \]  \hspace{1cm} (22)

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix} s_k \end{bmatrix} + \mu_k
\]

If hemorrhage was not found for several frames, the system recognizes that hemorrhage was stanched and stops tracking.

### 3.1.3. Result

The actual output of the hemorrhage recognition is displayed in Fig. 11. This also depicts that the proposed method outperformed the segmentation only technique, where the such method was failed to detect hemorrhage in some frames. When hemorrhage is recognized during the surgery, a red warning sign on the top left corner is initiated.
Fig. 11. Results of hemorrhage detection. Left column: Manually traced hemorrhage region. Right column: automatically detected hemorrhage region (square) by the proposed method in two different laparoscopy videos. (a) Accurately segmented by segmentation-only method (red boundary) and detected by the proposed technique (b) Manually undetectable, falsely segmented by segmentation-only method, but accurately detected by the proposed technique from previously obtained information (c) Not segmented by the segmentation-only method, but accurately detected by the proposed method (d) hemorrhage recognition warning initiated as a red circle on the top left corner of the surgical view

The validation of the algorithm was proven with robot-assisted laparoscopic surgery videos containing hemorrhage events which were cut in at approximately 20s time period. The identified hemorrhage region was marked at the center of mass, indicating the median of the detected boundary pixels. The ground truth data sets were prepared by manually inspecting and marking the hemorrhage region in each image for comparison to the results obtained by the proposed method. To evaluate the performance of the proposed method, the root mean square error (RMSE) between the estimate and reference center of the mass of the hemorrhage region was computed. Moreover, the blood flow was analyzed by measuring the area of the detected blood region in each frame by counting the pixels inside the hemorrhage region boundary.
The RMSE profiles for the proposed method and segmentation-only method for an approximately 20s time period is displayed in Fig. 12. Between the frame number 100 and 120, the segmentation-only method was not able to detect the blood region temporarily hidden by obstructing objects, resulting in maximum RMSE values.

![RMSE profiles of hemorrhage detection.](image)

**Fig. 12. RMSE profiles of hemorrhage detection.** RMSE profiles of the propose method in comparison with the segmentation-only method in hemorrhage detection. In both cases, centroids of the manually traced hemorrhage region were used as ground truth values.

The average RMSE values over the entire time period for the segmentation-only method and proposed method for the two videos used were 5.6, 0.9% the image size, and 3.6 pixels, 0.7% of the image size, respectively. The obtained segmentation-only
method output implied the quite small distance error. However, the FKF used in the proposed method produced a more stable RMSE profile even during the occlusion period. As a result, the new approach outperformed the segmentation-only technique, which failed to accurately detect hemorrhage locations mainly due to the environmental distortions such as smokes, camera motions, or organ occlusions. RMSE analysis demonstrated that the proposed method could provide surgeons robust information about any unexpected hemorrhage that may occur during surgeries.

Fig. 13 shows area variation analysis that classifies types of hemorrhage as flowing or stagnant. The overall scheme is provided by the red line, which was calculated using peaks of area increase and decrease. After several frames, this operation specified targets of danger from warning signs. Fig. 13(a) shows a slow increase in blood that could be allocated as a stagnant hemorrhage. Other than frames 40 to 80, where the blood area calculated by color detection was reduced due to smoke and occlusions, the overall peak area evaluation showed linearity. Because of small overall area variations, the small variation peaks could be considered as non-persistent bleeding, thus being classified as a non-warning target. Fig. 13(b) shows an acute hemorrhage that rapidly flows around the surgical view controlled by surgeons after situation awareness. The fast
increase in the first 50 frames describes a rapid blood flow due to unintended agitation. The linear decrease after frame 60 shows the treatment and removal of hemorrhage by surgical tools such as forceps and suction. As interpreted in Fig. 13, the calculated area analysis depicts the surgical events during the surgery. Therefore, the blood-flow profile can be used in actual surgery for surgeons to control the blood flow rapidly by warning them of the state of the blood flow before they become aware of the situation.

In a complicated environment such as robot-assisted laparoscopic surgery, in which surgeons experience limited visions through laparoscopes, this automatic hemorrhage recognition function will help surgeons rapidly handle emergency situations through fast state-recognition. The proposed method alarms surgeons when they miss the occurrence of hemorrhage that is caused by mishandling and accidental contact of surgical instrument.

Because of the simplicity of the proposed computational algorithm, the calculation time for processing was short enough to be applied in an on-line manner. Automatic hemorrhage detection and alarming will provide meaningful information to surgeons for performing more safety-assured surgeries and reducing overall surgery time. This method will also be
extended to other surgical state-recognition applications that feature location with spatial and temporal accuracies is significant.
Fig. 13. Hemorrhage flow classification by area variation analysis. Red line indicated linearity of hemorrhage flow; blue line indicates the calculated hemorrhage area (a) stagnant hemorrhage (b) hemorrhage flow following stanch
3.2 2D Surgical Instrument Tracking

The universal algorithm applied for formulation of surgical instrument tracking using surgical images was carried out in three steps: 1) instrument detection with metal properties, 2) instrument tracking using similarity measures of sequential images and an adaptive filter, and 3) a damage warning using a Euclidean distance calculation. Obtaining metal color information under a vision system was first computed to segment multiple surgical instruments hypothesizing that all surgical instruments consist of similar gray color characteristics. The inter-frame difference calculation in the time domain followed metal detection, and a KF was applied to suppress the faulty detection error during instrument tracking. Finally, a safety shield region was set to each detected instrument, and the Euclidean distance was calculated to provide instrument collision and tissue contact warnings to users.

3.2.1. Feature Extraction

Before the segmentation process, instruments and organs were classified into different categories using k-means clustering under LAB space using the fact that surgical instruments consist of metallic characteristics such as a gray color. After a
chain of image processing techniques including entropy filtering, canny edge detection, and morphological closing and dilating, the output was combined with the results from instrument motion subtraction, which was the movement computed by subtracting inter-frame images. The inter-frame images were calculated by subtracting previous frame from the current frame. The organs movement is much slower relative to the surgical instrument movement, thus instrument movement could be verified as depicted in Fig. 14. Finally, each instrument was labeled.
Fig. 14. Surgical instrument segmentation. Upper row: selected frame images of laparoscopic videos during robot-assisted surgery. Lower row: results of segmentation process for surgical instrument detection.
3.2.2. Instrument Tracking

The labels were assigned to the instruments as each was first detected. Once an instrument was identified, the center position of the instrument was saved and displayed to track each position as in Fig 15 (e). When an instrument was initially identified, the status of the instrument was set to "new", and a lifetime value of 20% was assigned. When the instrument was found continually in several consecutive frames, its status was changed to "updated", and the lifetime value was increased by 20% for each frame up to 100%. When the instrument was lost in the frame after being detected, the lifetime value decreased by 20% at each frame, and the instrument status was set to "lost." When the lifetime reached 0%, the instrument label was set to "invalid," and the tracking of the instrument stopped, considering it as out of view. The out of view status was used if the instrument position was previously located near the image boundaries and lost in after-frame. The instrument numbers 1, 2, and 3 were assigned to the center point location, top–left, top–right, and otherwise respectively. If the instrument status was not "invalid," the instrument segmented image from the first step was used as a reference during the second step, similarity measure.

Finally, the last step of the algorithm, KF was utilized to
compensate for the tracking failures. The three location parameters, previous instrument position, the output from segmentation and template matching process were provided to calculate the KF input. Using these three parameters, Euclidean distance between the instrument center point output from the segmentation and the template matching step in the current image and the previous location of the surgical instrument detected, if known, was measured. The measurement noise covariance R was continuously adjusted according to the result of the Euclidean distance calculation. In the implemented KF, 6 states and 4 measurement which includes the offset difference between the outputs from the segmentation and template matching method were applied. The offset was due to different localization calculation technique, for example the segmentation technique calculates the instrument position from the boundaries of the instruments, and template matching method localized the instrument from the rectangular window of the instrument image. The offset was automatically estimated and works as a damping process between two different measurements by including the offset in the state variable. The implemented process and measurement model which include the offset were represented in (23) and (24), respectively. In (23), $\Delta T$ denotes the sampling time. As a result, the universal method improved the accuracy by using a KF built on acquired knowledge from the previous instrument location inputs.
\[ s_{k+1} = A s_k + w_k \]  
\[ = \begin{bmatrix} 10 & AT & 0 & 0 \\ 01 & 0 & AT00 \\ 00 & 1 & 0 & 00 \\ 00 & 0 & 1 & 00 \\ 00 & 0 & 0 & 10 \\ 00 & 0 & 0 & 01 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ x_k' \\ y_k' \\ \delta x \\ \delta y \end{bmatrix} + w_k \] 
\[ z_k = H s_k + \mu_k \]  
\[ = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} s_k + \mu_k \]  

\( R \) is the measurement noise covariance which was set as \( 4 \times 4 \) identity matrix modified according to the measurement input value. Because the multiple positions were observed from the 2 sensors, the segmentation and the template matching measurements, the weight of the sensor measurements via the measurement noise must be tuned. \( \gamma_c \) in (25) denotes the measurement error from the color detection observation, and \( \gamma_t \) denotes the measurement error from the template matching observation.

\[ R = \begin{bmatrix} \gamma_{cx} & 0 & 0 & 0 \\ 0 & \gamma_{cy} & 0 & 0 \\ 0 & 0 & \gamma_{tx} & 0 \\ 0 & 0 & 0 & \gamma_{ty} \end{bmatrix} \]
After the identification of instrument positions two types of warnings could be provided: a collision danger warning, initiated when the instruments had possibility to collide with one another, and prohibited region danger warning initiated when the instruments had possibility to damage particular tissues that should not be perforated. Because the instrument boundary information, such as size, could be extracted from the segmentation process, a "danger region" was defined and calculated around the instrument tip region shown in Fig. 15(f); moreover Euclidean distance measured the distance between instruments to notify collision warning. For tissue perforation warning displayed in Fig. 15(g), an initial “prohibited region” was declared at the start of the surgery and kept being tracked using template matching algorithm. As well as this warning function, Euclidean distance of the instrument and the prohibited region are measured for tissue perforation warning.
Fig. 15. Process of surgical instrument danger warning. a) Original image. Three instruments are shown. b) Binary image from k-means clustering and color classification. White regions show detected instruments. c) Object movement using sequential image contrast was used to suppress noise. d) Detected instruments from color segmentation with computed convex bounding box added (green), manually drawn instrument boundary (red). e) Labeled instruments with their center points displayed. f) Collision warning when instruments entered a defined prohibited region; red rectangle depicts instrument tip.
3.2.3. Result

Multiple instruments could be automatically tracked without any manual guidance. Different environments with various image distortion factors including smoke, organ/tissue occlusions, and camera motion were randomly chosen and used in various tests. In Fig. 15(f) and Fig. 15(g), the warning markers were overlapped on instruments that enter dangerous proximities to other instruments and tissues.

The quantification analysis was computed via three instrument trajectories using different position identification methods: one with the first step of the universal algorithm, segmentation; the second step, addition of similarity measure; the third step, addition of KF. They were compared to the ground truth data analyzed using manual detection. This is depicted in Fig. 16 and Fig. 17 where Fig. 16 shows path trajectories in position in time domain, and Fig. 17 is displayed in image domain. As the figure shows, the final step of the algorithm was the closest to the identified ground truth value. Quantification was done through a RMSE analysis, where RMSE for instrument 1, 2, 3 resulted 39 pixels, 6% of the image size, 15 pixels, 2% of the image size, 74 pixels, 12% of the image size, respectively. The average RMSE was 42 pixels, which was about 7% of the image size. The proposed method outperformed other two methods, which
contained noise peaks and failed to accurately detect instrument locations.

In addition, to verify the accuracy of the proposed algorithm, sensitivity and specificity of each surgical instrument detection were compared with the ground truth location data is displayed in Table 2. Sum of four real robot-assisted laparoscopic videos, including da Vinci surgical videos, were used. The average sensitivity was 86% and specificity was 96%.

![Graphs showing instrument position over time](image)

(a)
Fig. 16. Validation of surgical instrument position trajectories in time domain. The plot shows $x$ and $y$ position of a) instrument 1 b) instrument 2 and c) instrument 3 as reported by manually traced values (dashed line), and the proposed tracking algorithm (solid line).
Fig. 17. Validation of surgical instrument position trajectories in spatial domain. (a) Proposed method validation using multiple instrument path trajectories in comparison with manually traced values. $x$, $y$ pixel positions of instruments 1, 2, 3. (b) Path position distance error of instrument 1 (c) Path position distance error of instrument 2 (d) Path position distance error of instrument 3.
Without using an additional hardware, the danger warning system through multiple instruments tracking was defined to be feasible. One of the implementation was the dependency of segmentation accuracy performance on instrument color and surface shape characteristics. For example, the holes on the irrigation instrument surface yielded low accuracy due to light reflection; however this outcome may be less relevant to accidental injuries such as grabbing or cutting tissues compared to other instruments.

With the proposed functionality in surgical robot, surgeons will be able to provide safer surgery with fast environment perception. Biocompatible issues from artificial markers such as specialized instruments and separated cameras or detectors, from the previous studies can be avoided through full automatic image guidance. This functionality will increase the efficiency of surgical procedures by providing additional safety functions and useful information to surgeons. This functionality and collision warning system can be further developed with stereovision to increase the accuracy and deliver volumetric information.
<table>
<thead>
<tr>
<th>Surgical Instrument 1</th>
<th></th>
<th>Surgical Instrument 2</th>
<th></th>
<th>Surgical Instrument 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Control</strong></td>
<td><strong>Proposed algorithm</strong></td>
<td><strong>Control</strong></td>
<td><strong>Proposed algorithm</strong></td>
<td><strong>Control</strong></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
</tbody>
</table>

- indicates non-occurrences.
3.3 3D Surgical Instrument Tracking

In order to provide surgeons better view of the surgical field, 3-dimensional (3D) analysis that includes the depth information of the surgical instrument is needed to fully prevent the surgical instruments from perforation of the organs. Using the universal blob analysis proposed, surgical instruments were classified from the organs in both left and right images. They were then proceeded to the next step of depth extraction using 3D reconstruction algorithm. For validation, simulation using cylinder objects that mimic surgical instruments were implemented using C++, including opencv library, environment. The process is demonstrated in Fig. 18.
Fig. 18. Block diagram of 3D depth extraction and simulation.
3.3.1 Feature Extraction

As the universal algorithm suggests, the feature extraction was achieved through kmeans clustering through LAB space color conversion to separate instruments with organs and a chain of image processing algorithms, such as erosion, dilation, and open calculations. With the kmeans clustered metal image, HSV space converted and threshold metal instrument segmentation were added to remove the illumination. The object motion analysis through instrument motion subtraction with inter-frame analysis was added to segment the instruments.

These segmentation processes were applied to stereo-images, both left and right image, to find the instruments in each image. Such processes were shown in Fig. 19. They were then compared for the next step, the depth extraction of the surgical instruments.
Fig. 19. Image processing step of 3D surgical instrument tracking. Left denotes the left image, and right denotes the right image from stereovision (a) original left and right image (b) kmeans clustered images (c) HSV analysis and noise elimination added from universal blob analysis (d) combined surgical instruments extracted from the left and the right images
3.3.2 Depth Extraction

Depth extraction of the surgical instruments are proceeded in three big steps: 1) corner point detection using harris corner point to detect the corresponding points between the left and the right image, 2) Procrustes analysis to normalize the feature scale, and 3) Delaunay triangulation to reform the feature in groups of triangles. To find the depth, we created a 3D triangular shape of the surgical instrument and extract the depth from the shape. The universal method is used

**Harris Corner Point Detection**

To find the match point between left and right images, block matching is usually used in stereovision algorithms. However, we chose to use the corner point detection algorithm and find tentative correspondence, comparing similarity of the corner neighborhood in the searching window. Among various corner point algorithms, Harris corner point detection has the advantage over other corner detection methods, such as block matching detection or feature point detection, by presenting a very strong characteristic in structural features. Harris corner point detection algorithm was first created in 1988 by C. Harris. This method uses the intensity gradient and \(2xI_1D\) convolution to find the
best candidate via non-max suppression and threshold. By the means of corner detection, the corner in literature represents the intersection of two edges. It represents a point where the directions of these two edges change. Hence, the gradient of the image in both directions had a high variation, which can be used to detect it.

This algorithm was applied to the given surgical instrument that was extracted from left and right surgical images and found the corner points of the instrument as shown in the Fig. 20. We set the number of corner points to have the sensitivity of 0.1 and quality level of 0.00001 which had an output close to the maximum corner points in the image. In our case, the image was the surgical instrument feature. The higher sensitivity of corner detection might yield less feature corner points which might lead to higher errors in finding corresponding points between left and right images. Hence, we decided to release as many corner points as we could get for larger number of input samples and discard the non-corresponding points using SSD matching.
Fig. 20. **Automatic Harris corner point detection.** Output from automatic Harris corner point detection applied to the surgical instrument feature from (a) the left image (b) the right image. Each column demonstrates each surgical instruments.

After extracting the intensity of the corner points in each image, SSD calculations were implemented to obtain the strongest corner feature match between two images, shown in the Fig. 21. The geometric threshold was given as the distance of 5, and the validated point locations after the feature matching were inserted into matrices.
Feature matching using SSD algorithm might result outliers due to a similar intensity of surgical instrument image. Therefore we removed outliers that had large distance differences between the left and the right image under the fact that the same feature point positions of surgical instruments in left and right images were close together. This process was implemented using geometric constraint via distance calculations. Utilizing the input of the match point locations extracted from feature matching via SSD, a pixel distance threshold of 1000 was applied to remove the large distanced outliers between two images. The result from discarding the outliers was shown in the Fig. 21(c). After the removal, the finalized geometric inliers were obtained for the next step, 3D reconstruction.
Fig. 21. Feature matching via geometric threshold constraint. (a) original left and right extracted feature overlapped image (b) corresponding point matching after corner point detection for each instrument (c) removal of outlier from matched corresponding point for each instrument using geometric threshold constraint.
Procrustes analysis

In order to create an accurate 3D version of surgical instruments using a group of triangles, normalization needed to be performed, possibly done by Procrustes analysis. Procrustes analysis is a form of statistical shape analysis used to examine the distribution of a set of shapes. The name came from the Greek mythologist who made his victims fit his bed either by stretching their limbs or cutting them off. Using translation, rotation, and uniformly scaling the object, the size of the objects and the placement in space are freely adjusted. If two objects have the shape with different scales, for example a circle with various radius after implementing Procrustes superimposing, the two shapes will be coincided.

From each geometrical shape, we calculated the mean shape and subtract it so that the centroid lied on (0,0). Then the mean shape was calculated from all inputs, geometrical shapes. For each shape $I$, parameters for the size, the rotation and the translation between the shape and the mean shape were set to minimum. The iteration is proceeded 2~3 times for error optimization.

To remove the translational components by calculating the mean shape using (26) with given $k$ points in two dimensions, each
point was translated so that their mean is translated to the origin.

\[ x = \frac{x_1 + x_2 + \cdots + x_k}{k}, \quad y = \frac{y_1 + y_2 + \cdots + y_k}{k} \]  
(26)

To normalize the scale, scaling the object so that the root mean square distance, which is a statistical measure of the object’s scale or size, for each object point from the origin became 1 using (27).

\[ s = \sqrt{\frac{(x_1 - \bar{x})^2 + (y_1 - \bar{y})^2 + \cdots (x_k - \bar{x})^2 + (y_k - \bar{y})^2}{k}} \]  
(27)

Then, the object size was normalized when the points were divided by the object’s initial scale.

The removal of rotational component was more complex than translational or scale component. By referencing one object orientation over another, the other object was needed to be rotated around the origin until optimum angle of rotation \( \theta \) was found via calculating the minimum SSD between the corresponding points. With given object points of \( ((x_1, y_1), \cdots), ((p_1, q_1), \cdots) \), a rotation angle \( \theta \) was represented as (28), where \( u, v \) equals the rotated point.
\[
\begin{bmatrix}
u \\
v
\end{bmatrix} = \begin{bmatrix}
\cos \theta p_1 - \sin \theta q_1 \\
\sin \theta p_1 + \cos \theta q_1
\end{bmatrix}
\] (28)

By taking derivatives of \((u_1 - x_1)^2 + (v_1 - y_1)^2 + \cdots\) with respect to \(\theta\) and made it equal to 0, \(\theta\) became (29).

\[
\theta = \tan^{-1} \left( \frac{\sum_{i=1}^{k} (p_i y_i - q_i x_i)}{\sum_{i=1}^{k} (p_i y_i + q_i x_i)} \right)
\] (29)

Procrustes analysis was used to obtain similar placement and size by removing translation, rotation, and scale factors of the corresponding corner points extracted, the geometric inliers. Using these points of the surgical instruments feature, the eight refined points from the normalization process were resulted as shown in the Fig 22. The detected instrument was normalized and superimposed after Procrustes analysis.
Fig. 22. Feature point normalization after Procrustes analysis.
Green and red points each display different feature. Left column: original corresponding point placement, right column: superimpose after Procrustes analysis (a) for instrument 1 (b) for instrument 2

Block factorization

Using the shapes of the refined points from Procrustes analysis, 3D reconstruction process was conducted using a sub-block factorization algorithm, based on Tomasi–Kanade factorization. Because of the rigidity of the instruments and stereo image
characteristics, where usually more than ten shapes are used for the 3D reconstruction [75], and temporal index are included, the algorithm could be simplified.

Using the shapes composed of the refined 2D points from Procrustes analysis, 3D shape was constructed using low rank constraints. The method hypothesizes that when a shape from 3D plane is projected on to 2D plane, 2D shapes can be achieved, and a matrix consisting of the 2D shapes can be factorized into a projection matrix and a 3D shape.

To calculate the factorization, first, we defined $S$ to be a shape that is $3 \times P$ matrix describing $P$ points. The shape of a specific configuration in a linear combination is described as (30), where for our algorithm number of shapes, $K$ equals 2, and $l_1 \cdots l_k$ denotes the weight of the 3D shape based on the 2D shape.

$$S = \sum_{i=1}^{K} l_i \cdot S_i \quad (30)$$

Under a scaled orthographic projection, the $P$ points of a configuration $S$ were projected into 2D image points, $(u_i, v_i)$, shown in (31). In (31), $T$ could be eliminated due to normalization from Procrustes analysis. Here, $R$ represents the first two rows of a 3D camera rotation matrix.
Based on the equation (31) and (32), the relationship between 3D shape and 2D shape could be rewritten as the (33). The settings were set according to our algorithm, composed of two shapes.

$$W = \begin{bmatrix} l_1^1 R^1 & \cdots & l_K^1 R^1 \\ l_1^2 R^2 & \cdots & l_K^2 R^2 \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$$  \hspace{1cm} (33)$$

where

$$W = \begin{bmatrix} u_1 & \cdots & u_p \\ v_1 & \cdots & v_p \\ u_1^2 & \cdots & u_p^2 \\ v_1^2 & \cdots & v_p^2 \end{bmatrix}$$  \hspace{1cm} (34)$$

To simplify, we coded $Q$ and $B$ to be (35) and (36) respectively and rewritten to (37).

$$Q = \begin{bmatrix} l_1 R & \cdots & l_K R \end{bmatrix}$$  \hspace{1cm} (35)$$
Since the surgical instruments are classified as rigid, rigid factorization with no deformation factor can be implemented. Thus, (33) shows that the matrix has rank constraint $r \leq 3$, and can be factored into 2 matrices. $Q$ contains for the pose $R$, and the configuration weights. By decomposing the tracking matrix, $W$ in (37) into $Q$ and $B$ using singular value decomposition (SVD) in (38), the shape becomes (39).

$$SVD = W^{2K \times P} = \hat{U} \cdot \hat{D} \cdot \hat{V}^T = Q^{2K \times 3} \cdot B^{3 \times P}$$ (38)

Then we transformed $Q$ into a projection matrix $R$ as shown in (37) which was be reordered to (40), and the coefficients using rank 1. Additional linear transformed matrix $A$ is multiplied for the ambiguity of the $Q$ and $B$ as in (41). By the definition of orthonormal property of $Q$, where multiplication of the self matrix extracts an identity matrix, $A$ can be calculated using SVD. Then 3D mean shape, $B$ could be estimated as in (42).

$$B = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \quad \text{where} \quad B = \begin{bmatrix} x_1^1 & \cdots & x_P^1 \\ y_1^1 & \cdots & y_P^1 \\ z_1^1 & \cdots & z_P^1 \\ x_1^2 & \cdots & x_P^2 \\ y_1^2 & \cdots & y_P^2 \\ z_1^2 & \cdots & z_P^2 \end{bmatrix}$$ (36)

$$W = Q \cdot B$$ (37)
\[ Q = \begin{bmatrix} l_1 r_1 & l_1 r_2 & l_1 r_3 & l_2 r_1 & l_2 r_2 & l_2 r_3 \\ l_1 r_4 & l_1 r_5 & l_1 r_6 & l_2 r_4 & l_2 r_5 & l_2 r_6 \end{bmatrix} \] (39)

\[ Q = \begin{bmatrix} l_1 r_1 & l_1 r_2 & l_1 r_3 & l_1 r_4 & l_1 r_5 & l_1 r_6 \\ l_2 r_1 & l_2 r_2 & l_2 r_3 & l_2 r_4 & l_2 r_5 & l_2 r_6 \end{bmatrix} = [l_1] \cdot [r_1 \ r_2 \ r_3 \ r_4 \ r_5 \ r_6] \] (40)

\[ W = \hat{Q} \cdot \hat{B} = (QA^{-1})(AB) \] (41)

\[ B = A^{-1} \hat{B} \] (42)

Delaunay triangulation

With the refined points, Delaunay triangulation was performed to reconstruct the structure into 3D shape of surgical instruments through a group of triangles. Delaunay triangulation is a method that created triangles utilizing a set of points on a plane. This method maximizes the minimum angles of all the triangles. Implementing this process to the refined points is shown in Fig. 23. The surgical instrument reconstructed model overlapped onto original surgery image shown in Fig. 23(b) could not fully eliminate the noise. The part that is not the instrument, but was also counted as the instrument, was achieved due to similar darkness characteristics of the instrument. However, this can be neglected since the depth of the noise part is defined as deeper than the main body of
surgical instrument. In Fig. 24, the positions of the refined points and their depths were demonstrated in the image window; $x$, $y$, and depth positions for the refined points are shown in column-wise. This representation was later used for simulation for validation.
Fig. 23. 3D reconstruction of surgical instrument via Delaunay triangulation. (a) original image (b) 3D reconstructed instrument model overlapped onto the original image (c) 3D reconstructed model front view (d) 3D reconstructed model top view (e) 3D reconstructed model on side view
Fig. 24. Depth extraction. Depth of the surgical instrument feature points calculated from Delaunay triangulation, the numbers dedicate x, y, depth positions of the corresponding points are shown in column-wise.
Depth extraction of surgical instruments were used by the methods that are usually used in 3D object reconstruction rather than common multiple-view geometry technique. Multiple-view geometry starts from the projective reconstruction, refines it to affine projection and arrives at metric reconstruction. At the point of projective reconstruction when we implemented the algorithm, the ambiguity yielded errors. To extensively state, the auto-calibration process for the reconstruction ensured a linear solution, but it was sensitive to the estimation of the plane at infinity and was numerically unstable, or required additional camera information. In order to validate multiple-view geometry method, plane at infinity needed to be calculated from the scene information of cheiral constraints, such as camera views and focal point. The scene information may had noise and may had scale factor instability problem, thus it was complicated to extract the structural features.

From the lightning and occlusion effect of the image, it was difficult to extract the corresponding features using block matching, and not enough features could be obtained, displayed in Fig 25. Even if features were extracted from the images, the interpolation of the block matched points would be calculated to define the whole surgical instrument depth. Such interpolation would slow down the computation time and not aid in simplicity.
Fig. 25. **Stereovision block matching disparity map.** (a) block matching of original surgery images via sum of absolute difference (b) block matching of original surgery images via sum of squared difference (c) block matching of object extracted images via sum of squared difference (d) inverse of depth map shown in graph type. Higher blue bar graph represents the deepest, red bar graph represents the highest.
3.3.3. SIMULATION

For the validation of the 3D surgical instrument position, simulation was completed with opengl in C++. By using cylinder-style objects which mimic the surgical instruments, the movement positions of the surgical instruments could be virtually simulated.

By employing the feature points obtained from Delaunay triangulation and a convex hull boundary calculation, the four boundary points - top left, top right, bottom left, and bottom right - and a centroid of the surgical instrument could be calculated. The anterior and posterior point of the cylinder could be calculated with these four boundary points, and the centroid could be obtained from trigonometry calculation.

Since the surgical instrument tracking was accomplished using MATLAB, the stored files including the position of the surgical instruments were read and simulated, shown in Fig. 26.
Fig. 26. Surgical instrument tracking simulation. (a) Original image. (b) Simulation using cylinder-style instrument.
Chapter 4

Discussion and Conclusion

In the robot-assisted surgery, surgeons often face difficulty handling vascular injury due to environmental perception. Hence, the proposed automatic object recognition warning function will help surgeons to handle emergency situations rapidly and reducing mortality rate caused by exsanguination during vascular injuries.

The significance of timely detection and management of incidental tissue perforation, where the signs via hemorrhage, or collision of instruments has recently arisen. Since most existing object recognition methods focus only on one moving object, we developed a novel core scheme for detection and tracking algorithm that works for both hemorrhage and surgical instrument in laparoscopic video images taken during real robot-assisted surgery. However, in extraction of hemorrhage feature, kmeans clustering technique was not implemented due to the similar intensity classification or organs. Thus, classification of hemorrhage using color information was used and provided better detection outcome. By extending our
proposed method, the extra information and fully automated control, such as self-awareness of tissue perforation via organ tracking can also be obtained.

The performance of the proposed universal algorithm was evaluated with actual video data from robot-assisted laparoscopic surgeries. For the validation analysis for hemorrhage detection, each step-wise method was compared and discussed. The segmentation-only method worked well providing average value of 5.6 pixels, which is about 0.7% of the image size. This result showed that the distance error was considerably small. However, this difference was further reduced to 3.6 pixels when the proposed universal method was applied. Flow of hemorrhage measurement also demonstrated its feasibility by depicting surgical situations, which displayed a linear increase when blood was flowing and a decrease when the hemorrhage was stanched. The computation time was approximately 1.1Hz, which was due to longer video reading time when using MATLAB. Computation time can be dramatically reduced when programmed in C.

In addition to hemorrhage detection, surgical instrument recognition showed feasibility as well, realizing simultaneous multiple instrument tracking without using additional hardware. The error rate for the proposed method on the localization of
surgical instruments yielded the lowest error, providing 7% RMSE of the image size. The average sensitivity was 86% and specificity was 96%.

To provide more information on depth of the surgical instrument so that the universal method can be one step closer to application in tissue perforation prohibition, depths of the surgical instruments were obtained using 3D reconstruction method. The validation was done through cylindrical simulation.

Overall, the proposed method provided more accurate detection results by using mathematical filters built on inputs of acquired knowledge from previous object locations. It outperformed the segmentation-only technique, which failed to accurately detect the hemorrhage and instrument locations. The universal method implemented in this paper, following the structure of feature segmentation, similarity measure, and optimal estimation, acts one of the best solution in creating a common framework of object recognition during robot-assisted surgery. This algorithm can work as a base of the object recognition system during robot-assisted surgery because the algorithm utilizes the surgical image in every aspects. Object color characteristic classification, object information extraction in spatial as well as temporal domain is included. For localization optimization, optimal estimation algorithm is also used. By understanding the
aspects of surgical images, the universal algorithm provides satisfactory object recognition system through object information extraction. Because the proposed method contains various kinds of image information analysis, such as spatial, temporal, intensity, color information as well as optimal estimation, this method can be used as a base framework for any object detection in surgical images. Thus, the method can be called as a universal method.

Analysis on the proposed universal method on the automatic hemorrhage recognition and the surgical instrument tracking systems showed feasibility. The method will be able to provide surgeons a robust information about object movement and unsafe situations that might occur during the surgery. This can further be improved to manipulate the surgical robot instruments to move heavily when the instruments enter the prohibited region or organs using image guided system. This entrance information can be recognized with the warning sign. Hence, surgical warnings of unsafe surgery through automatic detection of hemorrhage and instrument collision detection would not only provide a meaningful information to surgeons for safer surgeries and reducing overall surgery time, but also establish patients a safer surgical outcomes. Moreover, this technique could provide a baseline for collision handling in fully robotized surgical system in the near future.
REFERENCES


[27] Lallas C.D. and Davis J.W., Robotic surgery training with


[54] Stradler P., Dvoracek L., Vitasek P., and Matous P.,


[63] Kanav Kahol N.C.K., Vineeth N Balasubramanian,


국문초록

지능형 수술로봇 시스템을 위한 영상 기반 자동 상황 인지 기법에 관한 연구

류지원
서울대학교 대학원
협동과정 바이오 엔지니어링

최근 그 사용이 증가되고 있는 수술용 로봇 시스템에서 의도하지 않은 출혈 및 장기 손상은 치명적인 결과를 초래하므로 빠른 수술 상황 인지와 그에 따른 적절한 대처가 필요하다. 그러나 복강경 영상의 좁은 시야와 시술 부위에 대한 집중 때문에 수술자의 빠른 상황 판단이 어렵다. 따라서 응급상황 발생 시 수술자의 빠른 상황 인지를 돕기 위한 자동 사물인지 기반의 위험요소 감지 및 회피 기술 개발은 지능형 수술로봇 시스템 구현에 중요한 요소이다. 본 논문에서는 수술 영상에서 상황인지에 사용될 수 있는 대표적인 정보로서 출혈, 수술도구 위치, 그리고 수술도구의 3차원 깊이 측정 등에 보편적으로 사용될 수 있는 영상처리 기법을 제안하였다. 제안된 방법은 잘 알려진 영상 및 신호처리 기법인 물체 분할 (segmentation), 형판대응(template matching) 및 칼만필터(Kalman filter)를 조합한 새로운 복합 영상처리 기법이다. 색물과 형태학 정보를 이용하여 대상 물체를 분할 검출한 결과와 영상 프레임간의
대응 정보를 칼만 필터에 입력하여 최적의 위치 정보를 추정하였다. 개발된 방법의 성능평가로서 출혈 인지 기능에서는 평가자가 수동으로 입력한 위치와 제안된 방법으로 자동 검출된 출혈 위치 간의 표준오차(RMSE)를 통해 평가하였으며 출혈의 호름에 대해서도 분석하였다. 수술도구 위치 추적에 대한 성능평가 역시 수동적으로 입력된 수술도구의 궤적에 대한 자동 알고리즘의 민감도와 특이도를 통하여 평가하였다. 또한 수술도구의 3차원 깊이 정보를 추가하여 수술 중 도구 움직임을 실시간으로 시뮬레이션 함으로써 응용 가능성을 보여주었다. 출혈 측정에서 평균 표준오차는 이미지 크기(640×480)의 0.7%이었으며, 출혈양이 증가하거나 지혈이 된 경우 그에 따른 출혈 면적의 평균적인 추세선이 각각 증가 및 감소하는 것을 확인하였다. 수술도구 위치 측정의 평균 표준오차는 이미지의 7%, 수술도구 검출에 대한 민감도는 평균 86%이었으며 특이도는 95%를 보임으로써 우수한 인지율을 보였다. 이와 같이 수술 상황 자동 인지에 사용하기 위한 영상처리기법의 성능 평가를 통해 제안된 새로운 복합영상처리 기법이 수술 도중 자동 사물 검출 및 위치 추적을 위한 보편적인 알고리즘으로 향후 지능적 수술 로봇 시스템 구현에 유용하게 활용될 것으로 기대된다.

주요어: 출혈 검출, 물체 추적, 로봇 수술, 수술도구 추적, 3D 복원, 상황 인지

학번: 2010-31019