# Chapter 4

# Endoscopy

Authors: Sven Haase and Andreas Maier



This chapter points out the key aspects of minimally invasive surgery with particular focus on abdominal surgery using endoscopes. The comparison between minimally invasive and conventional open surgery is illustrated and several procedures are detailed. Moreover, this chapter introduces the term *Image Guidance* and its benefits. Finally, this leads to the motiva- tion why range imaging is of special interest for the next step of modern surgery. As endoscopes are more or less regular cameras that are inserted into the body, we also use this section to introduce the pinhole camera in Geek Box 4.1, elaborate on fundamental mathematical concepts of projec- tion in Geek Box 4.2, and introduce homogeneous coordinates and perspec-tive spaces in Geek Box 4.3. Fig. 4.1 displays an endoscopic image taken in the sigmoid colon which is the part of the large intestine that is closest to the rectum.

## 4.1 Minimally Invasive Surgery and Open Surgery

One criterion to categorize medical interventions is the degree of invasiveness. Therefore, we distinguish between minimally invasive and open surgery which is illustrated in Fig. 4.2. As the notation suggests, minimally invasive procedures describe medical procedures with little operative trauma. In com-

© The Author(s) 2018
A. Maier et al. (Eds.): Medical Imaging Systems, LNCS 11111, pp. 57–68, 2018.
https://doi.org/10.1007/978-3-319-96520-8_4

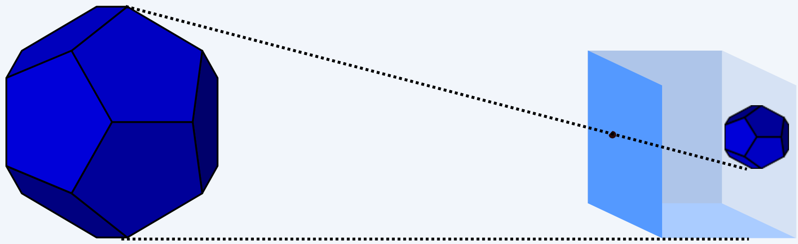

**Figure 4.1:** View into the colon using an endoscope. Here, the so-called sigmoid colon is under investigation for abdominal pain. The inner walls of the colon appear healthy without any signs of inflammation. Image data courtesy of University Hospital Erlangen, Germany.

parison to conventional open surgery, this leads to a shorter recovery time for the patient and thereby to a reduced hospital stay. In recent years, a variety of minimally invasive alternatives to conventional open surgery have evolved with special focus on pathologies of the heart and the abdomen. In contrast to open surgery, the physician has no direct access to the organs or structures of the human body. On the one hand, this means fewer and smaller scars and less pain for the patient, but on the other hand, without direct access the physician has a limited sense of orientation and usually has to rely on additional imaging techniques. For some medical procedures minimally invasive alternatives are not available as the incision is just too small, e. g., the removal of larger organs or transplantations. For smaller organs such as the kidney or gallbladder, laparoscopic interventions are already performed as a common routine. In terms of operative time, minimally invasive surgery usually takes longer due to the smaller incision and worse orientation. Both open and in most cases minimally invasive surgery require anesthesia during the intervention. Statistical comparison of open and minimally invasive surgery in terms of quality-of-life shows an overall improved result for laparoscopic interventions.

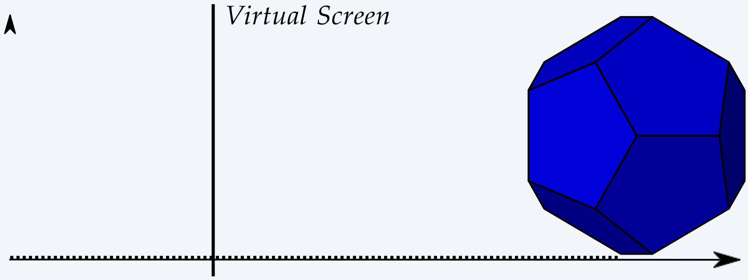
## 4.2 Minimally Invasive Abdominal Surgery

As one of the most important fields of minimally invasive procedures, the diagnosis and treatment of abdominal pathologies is the main medical application of this chapter. Here, a variety of important special instruments are required:

**Geek Box 4.1: Pinhole Camera**

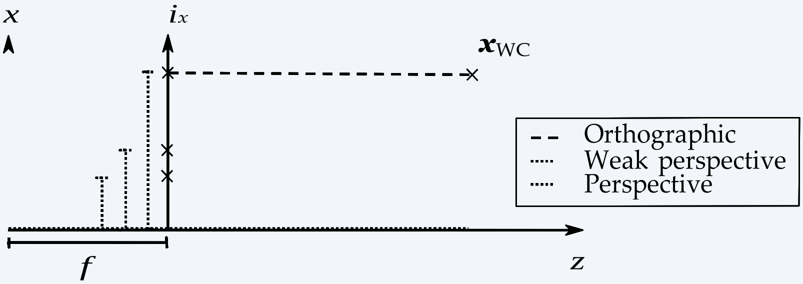


The pinhole camera model is a simple yet powerful model that allows us to describe various effects of the projection process. The model assumes the the outside world is observed using a pinhole. As such the outside world is projected onto a screen on the opposite side of the pinhole. Due to the pinhole, the image of the world is a scaled, upside-down version of the outside world. Doing so, the model allows us to describe perspective, i. e., that the size of the projection is dependent on the distance. In medical imaging and computer vision [2], the pinhole camera model is a widely used assumption. Another convention that is also commonly used is that the screen is virtually placed outside the camera for graphical simplification.




- *Endoscopes*: Depending on the procedure these devices are non-rigid (flexible endoscopes) or rigid (laparoscopes) and serve as a camera inside the human body. Rigid endoscopes have the benefit that the navigation is much more intuitive although the degrees of freedom during the navigation is reduced compared to non-rigid endoscopes.
- *Trocars*: To allow a fast exchange of different instruments a trocar is placed in the human body as a port to the abdominal cavity. For different procedures, different sizes ranging from several milimeters up to a few centimeters are used.

**Geek Box 4.2: Mathematical Projection Models**




In order to describe camera projections in math, we employ linear algebra and matrix calculus. The figure above compares the projection of a point $\boldsymbol{x}_{WC} = (x, y, z)^\top$ onto a virtual screen at coordinate $\boldsymbol{i} = (i_x, i_y)^\top$ at focal length $f$ (cf. Geek Box 4.1). For simplicity, we neglect the $y$ components in the figure.
The orthographic projection (dashed line) simply neglects the distance to the screen and finds the projection as

$$\begin{pmatrix} i_x \\ i_y \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Note that this model is not able to describe scaling as it occurs in projective modeling.
In order to alleviate this problem, the weak perspective model (line-dotted line) can be employed. Here we introduce a scalar value $k$

$$\begin{pmatrix} i_x \\ i_y \end{pmatrix} = \begin{pmatrix} k & 0 & 0 \\ 0 & k & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

that allows us to fix a global, depth independent scaling.
At this point, we observe that we are not able to find a linear model that is able to describe a full perspective projection model (dotted line)

$$\begin{pmatrix} i_x \\ i_y \end{pmatrix} = \begin{pmatrix} \frac{f \cdot x}{z} \\ \frac{f \cdot y}{z} \end{pmatrix} = \begin{pmatrix} ? & ? & ? \\ ? & ? & ? \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

for any focal length $f$.

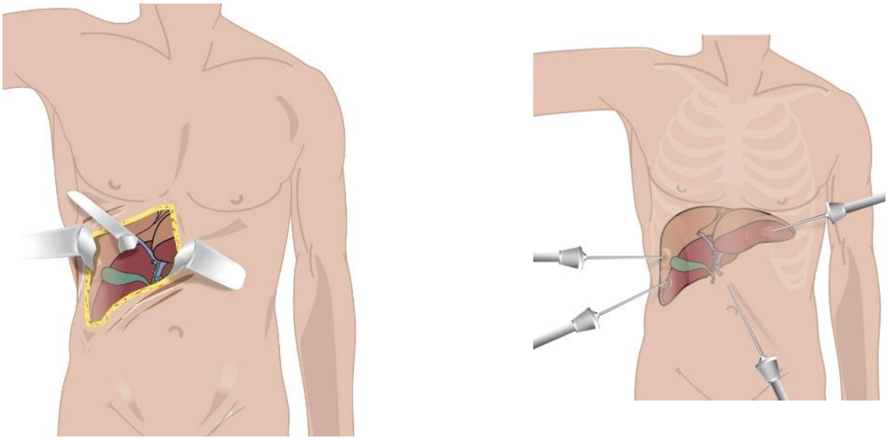


**Figure 4.2:** Illustrations of a cholecystectomy. The left image shows an open surgery with direct access and the right image shows a minimally invasive approach with endoscopic tools (Courtesy of Prof. Feußner, Technical University Munich, Germany).

- *Surgical instruments*: For the actual procedure, different endoscopic tools are required, e. g., clamps or scissors. Those instruments have a scissor-like grip to control the action at the top of the tool.

The workflow of an endoscopic procedure in general is described by four steps. Usually, the patient has to be anesthetized in a first step. The actual procedure starts with small incisions where the trocars are inserted. Then, the abdomen is insufflated with carbon dioxide gas. This allows the physician to have more room for the procedure. Finally, the endoscope and the tools are inserted through the trocars to start the actual treatment.

## 4.3 Assistance Systems

Modern surgery companies developed a variety of assistance systems to ease the navigation or to reduce the required manpower for minimally invasive procedures. Usually, several assistants besides the physician are required during the intervention. As the surgeon performs the actual procedure with endoscopic tools, one assistant has to hold the endoscope, one has to hand him the required instruments over, one has to supervise the patient and often a few more are involved for general organization. The remainder of this section introduces some of the available assistance systems in detail.

A very basic and intuitive assistance system is an endoscope holder. These medical devices are available in different complexities, ranging from simple

**Geek Box 4.3: Homogeneous Coordinates**

In Geek Box 4.2, we have seen that projective transforms cannot be expressed by means of linear transforms in a Cartesian space. However, there exists a class of mathematical spaces in which this is possible. In order to do so, we virtually extend the dimension of our vector $\boldsymbol{x}' = (x', y')^{\mathsf{T}}$ by an additional component. This way, we create a new *homogeneous* vector

$$\tilde{\boldsymbol{i}} = \begin{pmatrix} \lambda \cdot i_x \\ \lambda \cdot i_y \\ \lambda \end{pmatrix}$$

with $\lambda = 0$. Note that $\boldsymbol{i}$ can easily be lifted to homogeneous space by selecting, e. g. $\lambda = 1$. For conversion back to a Cartesian space, one selects simply the last component of a homogenous vector and divides all components by it. Then, the last component can be omitted and the point can be mapped back to the original space. This *projective space* has a number of additional properties, e. g., all points are only equal (denoted by $\cong$), if they map back to the original Cartesian point and points with 0 in the last component cannot easily be mapped back to the original space as they lie at *infinity*. A very good introduction is given in [2].

Using this powerful concept, we can now return to our previous problem and observe

$$\boldsymbol{i} = \begin{pmatrix} i_x \\ i_y \end{pmatrix} = \begin{pmatrix} \frac{f \cdot x}{z} \\ \frac{f \cdot y}{z} \end{pmatrix} \cong \begin{pmatrix} f \cdot x \\ f \cdot y \\ z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

non-electronic static holders to automatic flexible holders that are navigated with a joystick. In general, these endoscope holders allow a more stable image acquisition as they exclude any jitter induced by a human. However, those systems are often very basic and still have to be navigated by a surgeon, e. g., the *SOLOASSIST*, an electronic assistance arm that is navigated by a joystick and simulates a human arm.

Besides endoscope holders, fully automatic robotic assistance systems are already commercially available. These systems allow the surgeon to be at a separate workstation as illustrated in Fig. 4.3. All commands are directly transmitted to the robot allowing the surgeon to even be at a distant place while performing the procedure. One of the most wide-spread is the *da Vinci* system. It is navigated by grips and pedals that enable various degrees of freedom. For intuitive visualization, this robot acquires stereoscopic images and thereby gives the surgeon a 3-D impression of the scene.

**Figure 4.3:** The da Vinci assistance system (©2014 Intuitive Surgical, Inc) with the workstation on the left and the distant robotic system on the right.

## 4.4 Range Imaging in Abdominal Surgery

Although assistance systems for minimally invasive surgery are commercially available, one major problem of endoscopic interventions still remains. The orientation and thereby the navigation within the human body depends on the experience of the surgeon. Due to the lack of intuitive visual comparison to the environment, the narrow field of view induces a loss of depth and size estimation in the abdominal cavity. However, this information is required for diagnosis, e. g., the size of a polyp, and for decision making, e. g., choosing the most reasonable endoscopic instrument. Therefore, a variety of different approaches to compensate for this loss of information were investigated, e. g., adding grids to the endoscope lens or estimating sizes by comparison with known instruments. To achieve more accurate estimations, further approaches considering range images have been investigated. Although these systems can also be utilized to measure sizes or distances, range data acquiring devices also enable completely new projects to assist minimally invasive procedures, e. g., augmenting 3-D range images with preoperative 3-D CT data. Several different techniques have been investigated to acquire endoscopic range images. Besides several *Shape-from-X* approaches, e. g., Shape-from-Shading, the three most popular acquisition techniques are using *stereo vision* setups, *structured light* or time-of-flight (TOF). Today, only stereo endoscopes are commercially available, but the two other techniques are highly investigated by different researchers. This section will focus on the working principle of available range imaging systems.

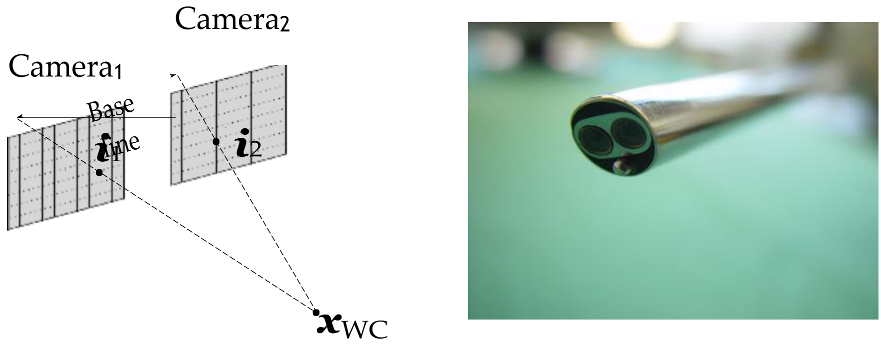



**Figure 4.4:** The illustration on the left hand side describes a stereo vision setup with two cameras $\text{Camera}_1$ and $\text{Camera}_2$ that observe a 3-D point $\boldsymbol{x}_{\text{WC}}$. The projection of this point onto each image plane results in a pixel index $\boldsymbol{i}_1$ and $\boldsymbol{i}_2$ which can be computed by a simple pinhole camera model. The right hand side image shows the front view of a stereo endoscope with two apertures to acquire images from two different field of views (close-up view courtesy by Prof. Speidel, National Center for Tumor Diseases, Dresden)).

## 4.4.1 Stereo Vision

Stereo endoscopes are the most investigated range image acquiring setups in minimally invasive surgery. The principle is also implemented in the da Vinci assistance system. Stereo vision describes an intuitive acquisition technique that is similar to the human vision and depth estimation.

The core concept behind stereo endoscopy is to estimate range information by observing a scene from two different perspectives. Given a known baseline, the framework has to detect the 2-D projections of a 3-D point in both image planes. In theory, using basic trigonometry, the range information of these points can then be calculated by triangulation, see Fig. 4.4. In practice, both lines will probably not intersect and minimizing the distance of both lines will estimate the position of the 3-D point.

The requirements for stereo endoscopy are on the one hand a precisely calibrated device and on the other hand diversified texture information of the observed scene. Accuracy is increased with a wider baseline between both sensors. As this baseline is limited by the diameter of the endoscope, the improved accuracy has to be gained by calculating the corresponding points in both images with higher precision. Corresponding points are calculated by detecting features in both images, e. g., by applying the scale-invariant fea- ture transform (SIFT) or by computing speeded-up robust features (SURF). Matching those feature points results in point pairs that correspond to the same 3-D point in the observed scene. Therefore, the output of a stereo endoscope highly depends on the quality of the two images and on the speed and robustness of the feature detection and matching. The estimated range images of a stereo setup, also called disparity maps, have a scene dependent image resolution.

The bottleneck of this range image acquisition technique is the calculation of corresponding feature points. Hardware manufacturers tackle this problem by increasing image resolution in the sensor domain. This leads to more details even in almost homogeneous regions in the acquired images, but also induces more computational effort to calculate features on both images. Therefore, estimating accurate 3-D range data in real-time is a major issue in stereo endoscopy. Stereo endoscopes are currently the only commercially available and CE certified 3-D endoscopes.

### 4.4.2 Structured Light

Structured light endoscopy is a novel technique based on stereo vision concepts, but with artificially created feature points instead of those given by textural information. In minimally invasive surgery structured light systems are not yet commercially available.

The working principle of structured light sensors is very similar to stereo vision systems. In comparison to those, structured light systems do not require two cameras observing the scene. The second camera is replaced by a projector that generates a known pattern onto the observed scenario, see Fig. 4.5. The known baseline and corresponding points in the acquired image and the known projection pattern are used to reconstruct 3-D points by triangulation. As the projector generates the pattern that is used to calculate the feature points, structured light is also called an active triangulation technique.

Similar to stereo vision, the baseline between the sensor and the projector and an accurate calibration is required for high quality measurements. As long as the pattern is clearly visible, this technique is independent of texture information of the observed scene. In comparison to stereo vision systems, the feature detection framework can be highly adapted to the projection pattern, as the structure of the feature points is known. The projection pattern should be easy to detect and hard to disturb by the texture of the scene. In conventional structured light setups, stripes or sinusoidal patterns are popular.

As the core concept for structured light is similar to stereo vision, so is its bottleneck of detecting and identifying the feature points of the projected pattern. Furthermore, the smaller the structures of the projected pattern are, the more 3-D points can be reconstructed, but the harder it is to identify those structures in the acquired images.

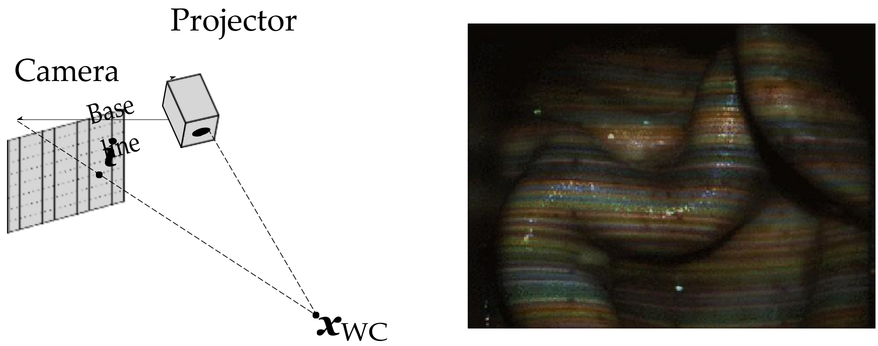



**Figure 4.5:** The illustration on the left hand side describes a structured light setup with one camera and one projector that generates a known pattern onto the scene. The projection of this point onto the image plane results in a pixel index $\boldsymbol{i}$. Together with the projection pattern the 3-D world $\boldsymbol{x}_{\mathrm{WC}}$ can be reconstructed using triangulation. The right hand side image shows a projection on animal organs.

### 4.4.3 Time-of-Flight (TOF)

TOF technology [4] tackles the topic of 3-D reconstruction from a completely different field of view. Instead of using acquired color images to find any distinctive structures, reflection characteristics are exploited to physically measure the distances of the observed scene. The first work to introduce this concept is found in [6].

The concept behind TOF technology is to measure a frequency modulated light ray that is sent out by an illumination unit and received by a TOF sensor, see Fig. 4.6. The received sinusoidal signal is sampled at four timestamps to estimate the phase shift $\Phi$ between the emitted and the received signal. The radial distance $d$ is then computed by:

$$d = \frac{c}{2f_{\mathrm{mod}}} \cdot \frac{\Phi}{2\pi}, \tag{4.1}$$

where $c$ denotes the speed of light and $f_{\mathrm{mod}}$ the modulation frequency.

As the illumination unit can be realized by an light-emitting diode (LED) and the sensor is a simple CMOS or CCD chip, production costs of TOF sensors are rather low. However, due to their novelty compared to stereo vision, current TOF devices exhibit low data quality and low image resolution. Besides the range image, most TOF devices provide additional data, e. g., photometric data, often denoted as the amplitude image and a binary validity mask. Due to its measurement technique, TOF setups do not require

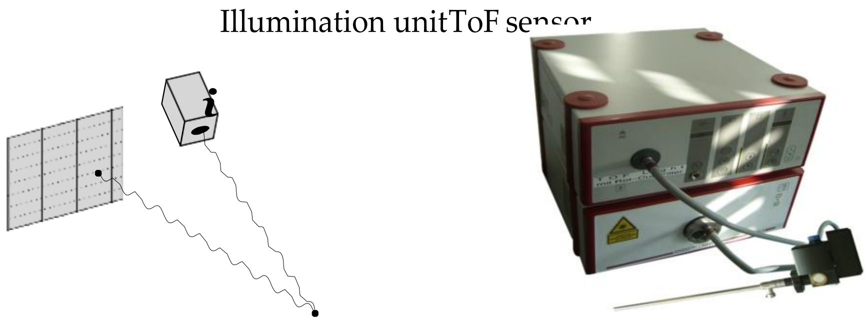



**Figure 4.6:** The left hand side illustration describes a TOF setup with a single camera and an illumination unit that sends out the modulated light. The reflected signal is then received by the TOF sensor at pixel index ***i***. Aftercalculating the phase shift, the radial distance is computed. The right handside image shows a prototype. The setup includes two separated light sources for the color and the TOF acquisition.

a baseline between the illumination unit and the measuring sensor, which isbeneficial for the use in minimally invasive surgery.

Besides the low resolution, systematic errors reduce the data quality ex- tremely. The error sources range from color dependent range measurements over temperature issues of the devices to flying pixels at object boundaries.In minimally invasive procedures, two major issues occur. First, multiple re- flection within the abdominal cavity corrupts TOF measurements. Second, inhomogeneous illumination caused by the endoscopic optic hinders accurate range measurements. Still the technology is real-time capable and thereby also suited for live broadcasting [9].