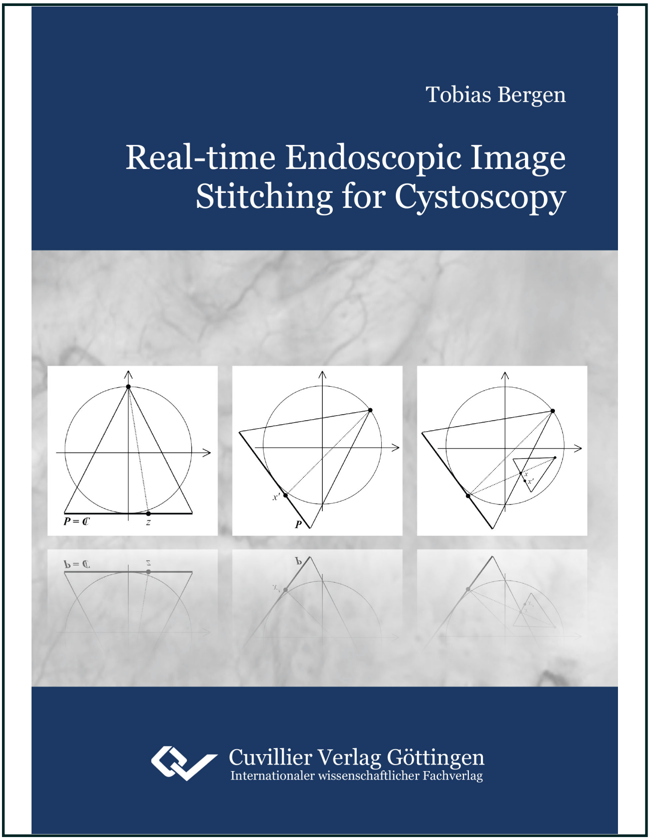
Tobias Bergen (Autor)

Real-time Endoscopic Image Stitching for Cystoscopy



https://cuvillier.de/de/shop/publications/7643

Chapter 1

Introduction

1.1 Improving Orientation and Documentation of Cystoscopies through Real-time Image Stitching

[1]Endoscopy is an established procedure for diagnosing and treating a wide variety of diseases and injuries inside the human body. The term *endoscopy* stems from the greek words *endos* (gr.: inside) and *skopein* (gr.: to look at). Medical endoscopes enable us to look into visceral cavities. They are used to inspect a variety of organs. Examples are the lungs and airways (bronchoscopy); the bladder, urethra, and ureter (cystoscopy); the stomach and esophagus (gastroscopy); the colon (colonoscopy); joint cavities (arthroscopy); and ear, nose and throat (sinuscopy, laryngoscopy). Furthermore, minimally invasive surgery (MIS) using endoscopic instrumentation is carried out e.g. in the abdomen, the brain, and on joints. Endoscopes are inserted into the body either through natural orifices, like mouth, nose, esophagus, airways, rectum, and urethra or through small incisions through the abdominal wall using trocars. [Kra11]

The advantages for the patient of minimally invasive inspections and procedures of this type are usually associated with reduced surgical trauma and shorter convalescence times. On the other hand, the techniques involved require a high degree of orientation, coordination, and fine motor skills on the part of the surgeon, due to the very limited field of view provided by the endoscope and the lack of relation between the orientation of the image and the physical environment (horizon). As a consequence, investigations into ways of providing computer assistance for endoscopic procedures have become a very active field of research in recent years. To overcome the problem of the limited field of view, stitching technologies based on image processing have been developed. The terms image stitching and mosaicking refer to the process of combining several (partially overlapping) images to create a broader field of view or panorama image with a wider perspective. The terms *stitching* and *mosaicking* are used synonymously throughout this thesis; the result of the process is referred to as *panorama image* or *image mosaic*.

---

[1] Parts of this chapter have been published in [BW16] © 2016 IEEE.

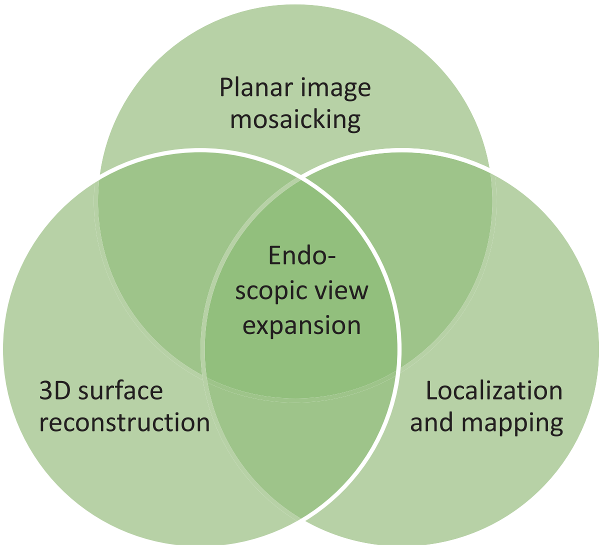


Figure 1.1: Endoscopic view expansion incorporates methods from different branches of research on image processing and computer vision. Approaches range from planar image mosaicking to 3D surface reconstruction as well as localization and mapping algorithms.

Classical image mosaicking, originally developed for stitching together sets of photographs, usually uses a planar surface which the images are project onto. If, however, the three-dimensional shape of the scene observed in the images significantly deviates from a plane and the camera performs any translational motion, the planar projection becomes erroneous. This is often the case in endoscopic applications, and generating a panorama image therefore also implies estimating the underlying surface geometry and camera motion. At this point, image mosaicking intersects with the field of three-dimensional (3D) surface reconstruction from a series of images, which has also been widely studied in the computer vision community. A further challenge emerges from the fact that the position of the endoscope is generally unknown. In robotics, a similar problem has been investigated. Determining the position of a robot within an initially unknown scene purely from the robot's own sensor information requires building a map of the environment and at the same time locating the robot within that map. This problem has become generally known as *simultaneous localization and mapping* (SLAM) and is also related to surface reconstruction and image mosaicking, if the robot's sensor is assumed to be a camera. Different methods of *image mosaicking*, *3D surface reconstruction* and *SLAM* have been applied to endoscopic image sequences with the goal of assisting the endoscopist by creating an enhanced field of view, often referred to as *endoscopic view expansion* or *dynamic view enhancement* (Fig. 1.1). Endoscopic view expansion may incorporate methods from classical image mosaicking approaches, 3D surface reconstructions techniques, as well as SLAM algorithms. For the sake of simplicity and in accordance with the common terminology in literature, the term *image stitching* is used in this thesis to describe any form of view expansion, independent of the underlying shapes (two- or three-dimensional) and methods (mosaicking, surface reconstruction, or SLAM).

Figure 1.2: Image mosaic of part of the urinary bladder. The gap indicates that parts of the tissue are not being inspected completely.

Endoscopic view expansion serves two purposes: improved orientation during the procedure and documentation after the procedure. *Orientation* can be facilitated for the surgeon by expanding the field of view as an online process during the inspection of the organ. That way, the surgeon can relate the image content of his current view to its context in order to be geared to anatomical landmarks. For *documentation* purposes a *panorama image* or *map* of the inspected organ (e.g. bladder wall) can be included into the electronic health record, on the one hand for later review in potential follow up inspections, one the other hand to facilitate communication among clinicians about a certain case. Furthermore, such a map can serve to improve the quality of an inspection by visually indicating its completeness. Any white spot in the map marks a potentially missed part of the organ. Fig. 1.2 depicts an image mosaic of the urinary bladder with parts not being inspected completely.

The aim of this thesis is the development of a *real-time endoscopic image stitching algorithm* for endoscopic view enhancement which is suitable to serve the two goals of dynamically extending the field of view for an improved orientation and creating a panorama image for documentation. The considered field of application is cystoscopy, i.e. the endoscopic inspection of the urinary bladder via the urethra. Cystoscopic image sequences, recorded in the operating room, serve as source of image data for design considerations, algorithm development, and evaluation throughout the thesis.

The thesis is structured as follows. After an introduction to the medical and technical aspects involved in the topic (this chapter), an overview of the current *state of the art* in endoscopic image stitching is presented in Chap. 2. Chap. 3 provides an *outline* of the algorithm developed within this thesis. A new *polynomial filtering*

Figure 1.3: A rigid cystoscope and a flexible video-laryngoscope.

approach is presented in Chap. 4. Based on this method, an algorithm is developed to *compensate for shading effects* in endoscopic images (Chap. 5). Different *motion and scene models* (planar and spherical) are considered in Chap. 6. A new motion model is presented, tailored to the application of image stitching of a spherical scene. Chap. 7 deals with the topic of *image registration*, probably the most crucial aspect of image stitching. The newly developed feature-based registration method, called *difference of paraboloids* (DOP), greatly benefits from applying polynomial filtering. The conjunction of all algorithmic components within a *multi-threaded stitching framework* is described in Chap. 8, followed by a presentation and discussion of the *results* obtained (Chap. 9) and concluding remarks in Chap. 10.

1.2 Endoscopy

First historical reports of instruments used to view inside the human body trace back to ancient egypt, where first proctoscopies (rectum examinations) have been performed. Technical advances have been made since the early 19th century with the development of a first primitive endoscope, called the “Lichtleiter” (light conducter) by Philipp Bozzini. Based on the development of the incandescent light bulb by Thomas Alva Edison in 1879, Maximilian Nitze and Josef Leiter developed a first rigid endoscope with an electric light source, the “Nitze-Leiter cystoscope”. The image was transmitted through a system of multiple lenses from the tip (*distal end*) of the endoscope to the eyepiece (*proximal end*). In 1960, Herold H. Hopkins introduced rod lenses to built a rigid endoscope, which greatly improved the image quality compared to the Nitze-Leiter cystoscope. This rod lens setup is still applied in modern rigid endoscopes. In 1958, Basil J. Hirschowitz invented a *flexible* gastroscope using fibers to transmit the light. Today’s endoscopes can still be categorized into flexible and rigid endoscopes. The main difference is the way of transmitting the light from the tip to the eyepiece. Using fibers as light conductors

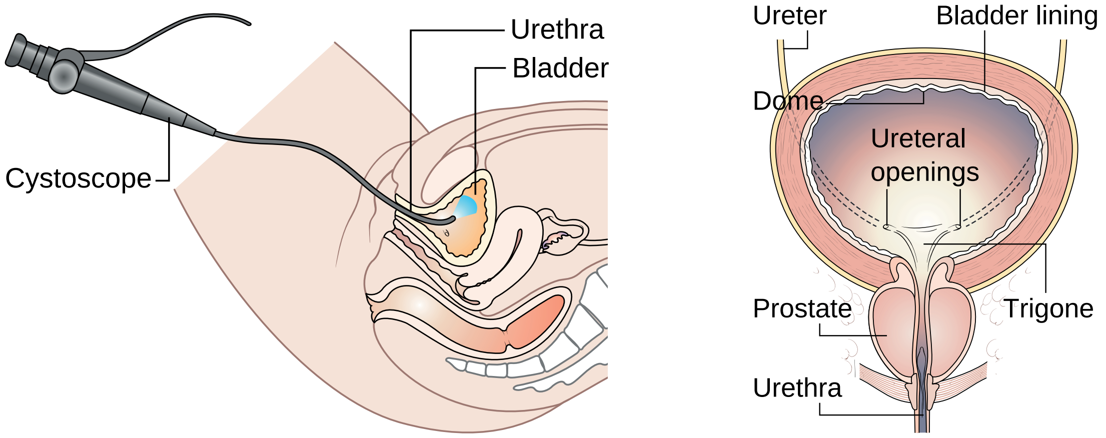


Figure 1.4: *Left:* Diagram of a cystoscopy with a flexible cystoscope on a female bladder [UK14b]. *Right:* Diagram of a male bladder [UK14a].

allows flexible endoscopes to bend when being inserted into the body, while rigid endoscopes transmit the light via a lens system built into a rigid shaft. Generally, the latter produce superior image quality, since the fiber technology limits the image resolution by the number of fibers used. To overcome this problem, modern flexible videoscopes, where a camera captures the images and displays them on a screen, are also available with the camera chip built into the distal end of the endoscope ("chip-on-the-tip"-technology). Fig. 1.3 shows a rigid and flexible endoscope. Rigid endoscopes are built with different viewing directions and different viewing angles. Available viewing directions range from 0° to 120° with viewing angles between 70° and 140°. Besides the endoscope itself, modern endoscopic systems usually consist of a light source and a digital camera system to capture the images and display them on a screen. [Wit14, Kra11]

1.3 Cystoscopy

Cancer of the urinary bladder is the 7th most common cancer worldwide. About 260,000 new cases are estimated to occur each year in men, and about 76,000 in women [EWI06]. The gold standard for diagnosing patients with bladder cancer is white light cystoscopy (in combination with urine cytology) [ST15, BBZ$^+$13]. Cystoscopy is the visual examination of the urinary bladder via the urethra. Both, flexible and rigid cystoscopes can be used. While flexible cystoscopy is more comfortable for the patient, rigid cystoscopes offer the advantage of an additional channel to insert endoscopic instruments to take biopsies or resect tumors. Rigid cystoscopy is usually performed under general anesthesia. A lubricant is used during insertion of the cystoscope through the urethra and the bladder is filled with a sterile fluid for the examination [MHH09]. A constant rate of flow keeps the size and shape of the bladder stable. Fig. 1.4 (left) shows a diagram of a cystoscopy using a flexible cystoscope on a female bladder. After gently inserting the cystsoscope through the urethra, the lining of the bladder wall is carefully examined by maneuvering the tip of the scope over all parts of the bladder wall in a systematic manner. The shape

1. Urachus
2. Left wall
3. Left ureteral orifice
4. Neck
5. Prostate
6. Trigone
7. Right wall
8. Posterior wall
9. Dome

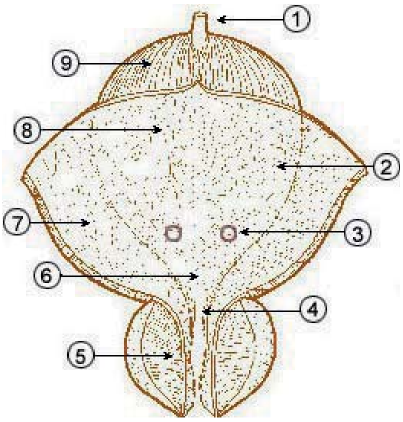


Figure 1.5: A diagram of the bladder is suggested to be used for documenting the position of possible findings [NCI].

of the bladder depends on its filling level. In a moderately filled state its shape is described as oval (ref. Fig. 1.4, right) [Cor46, p. 538].

The guidelines of the European Association of Urology state that “cystoscopy should describe all macroscopic features of the tumour (site, size, number and appearance) and mucosal abnormalities. A bladder diagram is recommended.” [BBZ$^+$13, p. 35] Nonetheless, in clinical practice the form of documentation varies greatly. Often, documentation is performed in written form only, potentially complemented by a diagram to further describe the location of a finding (ref. Fig. 1.5). Video cystoscopes allow to further capture still images or videos during the examination. While still images help to document relevant findings pictorially, they lack the information of where a captured region is located within the organ. Including video sequences into the documentation, which display the findings in spacial relation to anatomical landmarks, may help to overcome this problem. However, the fact that watching a video of a previous examination is a time consuming process raises doubts if this is practicable within the clinical work flow. Panorama images of relevant parts of the organ surface have the potential to greatly increase the value of findings documentation. They enable a surgeon to create a larger view which shows a finding in relation to natural landmarks, like ureteral openings, the trigone, and the dome. Fig. 1.6 shows a panorama image of the left bladder wall with a tumor located above the left ureteral opening. For documentation purposes it is not necessary to capture the entire bladder wall. The stitching algorithm should be able to create panorama images that are large enough to capture the clinical findings and relevant anatomical landmarks.

Regarding the task of cystoscopic image stitching within this thesis, the following assumptions are made:

Duration: The complete examination of the bladder is performed within a few minutes. Cystoscopic recordings available for this thesis have a duration between 8 and 70 seconds.

Shape of bladder: If the filling level of the bladder is kept constant during the examination its shape can be approximated by a sphere. The organ is as-

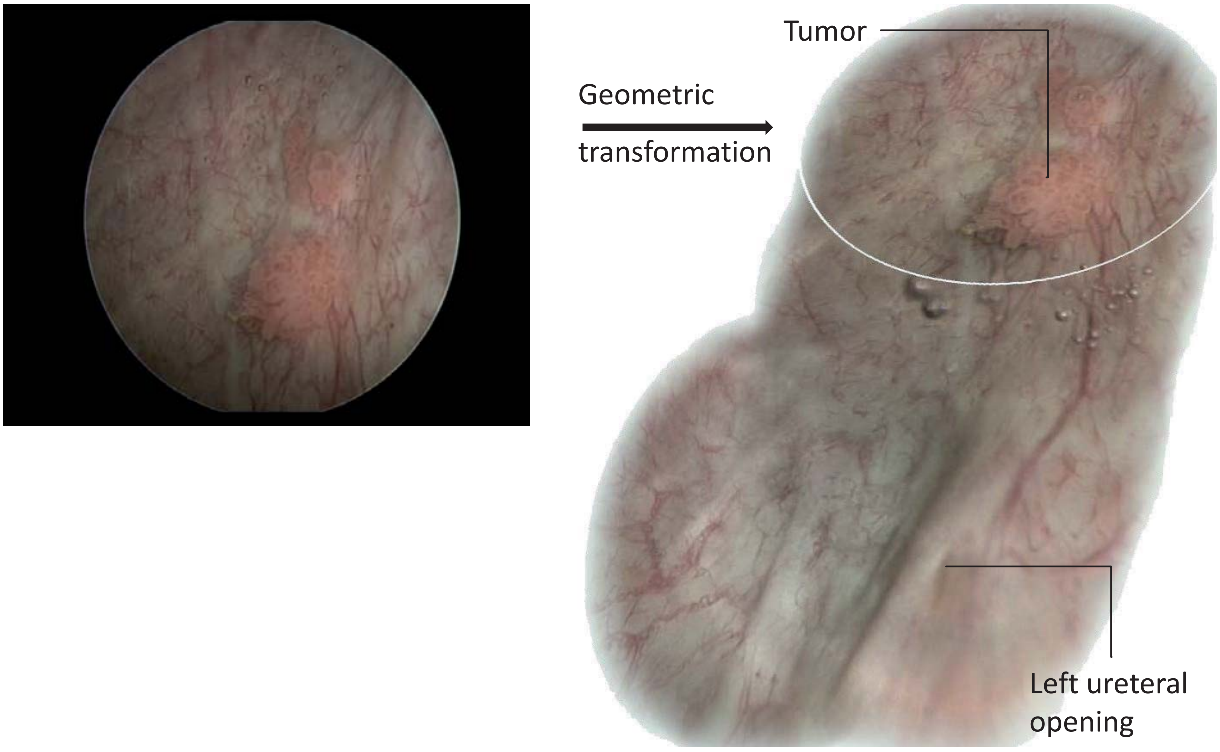


Figure 1.6: *Left:* Still images of a finding can be added to the documentation of a cystoscopy. *Right:* A panorama image of the bladder allows to spatially relate the tumor to natural landmarks, such as ureteral openings. The white circle marks the field of view of the still image on the left.

sumed to be rigid, so that no perceivable deformation takes place during the examination.

Size of bladder: The capacity of the urinary bladder varies greatly among individuals between 40 and 510 ml. Nonetheless, when filling the bladder during cystoscopy, a volume of 200 to 300 ml is recommended, resulting in a surface area of 165 to 217cm^2 (assuming the bladder has a spherical geometry). [Cor46, p. 538]

Field of view: The expected area A of a single field of view can be estimated by $A = \pi(d \tan \frac{\alpha}{2})^2$ with d being the distance of the tip of the cystoscope to the bladder wall and α the endoscope's viewing angle. The typical viewing angle of a cystoscope is around 90°. Assuming a typical distance d between 0.5 and 2 cm, the field of view has a size of 0.78 to 12.5 cm^2.

Number of images: To estimate the expected number of images, which have to be captured to cover one hemisphere of the bladder, the surface area is divided by the size of one field of view. If an overlap of 80% is required between consecutive images, between 35 and 720 images can be expected to contribute to the final panorama image.

1.4 Challenges and Requirements

Developing an image stitching algorithm for endoscopic images poses special challenges. These will be structured into three categories, according to their relation to 1) the endoscopic device, 2) the scene, and 3) the handling by the surgeon.

First, the technical characteristics of endoscopic image acquisition are considered. Image resolution is often quite low. Typical resolutions vary from 720 × 576 pixels (PAL) to 1920 × 1080 pixels (Full HD). Usually, only part of the image contains relevant information due to the aperture mask. This effect can be observed in Fig. 1.6 (left). Clinical recording systems often reduce the resolution further and apply compression algorithms to reduce the amount of data that has to be stored. Illumination can be very inhomogeneous, as the light source in the endoscope is focused and aligned with the optical axis, leading to a decrease in brightness from the center of the image towards the edges. In addition, most endoscopes are set up with wide-angle lenses, causing major geometric image distortions.

Second, the condition of the scene being observed poses certain challenges. The tissue being examined often has little or poorly contrasted texture; surgical instruments may be moving within the field of view; and staining from blood or other body fluids make robust image processing difficult. A major challenge can also arise from deformation of tissue, which violates the assumption of a rigid scene, implicitly made in many algorithms.

Third, the handling of the endoscope by the surgeon is generally not constrained, usually leading to three rotational and three translational degrees of freedom for camera motion. If the device is moved quickly, motion blur is commonly observed in endoscopic images.

Besides all of these challenges, also the requirements arising from endoscopic applications need to be considered, which in some respects may be less demanding than in other fields of application. Stitching photographs requires highly accurate registration results, because even small misalignments cause irritating artifacts and disturb the esthetic impression of the panorama image. In endoscopic image stitching, usually compromises can be made in terms of the accuracy of the reconstruction or texturization of the scene, so that in contrast to photographic image stitching, some visually perceivable deficiencies may often be acceptable, as long as they do not have any impact on the clinical interpretation of the image. Depending on the application, computation must be performed in real time, or may last up to several hours. For online view expansion, the processing speed is crucial and image mosaicking has to be performed in real time. On the other hand, if panorama images are being generated for *documentation* and *quality assurance* only, the computation time is less of an issue. An important requirement arises from the fact that interference of the system with the surgical routine has to be kept to a minimum. Visualization enhancement through image stitching should assist the surgeon, not vice versa. This restricts the possibility of user interaction and prohibits special demands on the clinical protocol in order to assist the system.

For the development of a stitching algorithm for cystoscopic image sequences within this thesis, the following prerequisites and requirements are identified:

Input: The input consists of video recordings of cystoscopic examinations. The recording equipment, available in the operating room, is used. Thus, the video sequences are encoded as compressed MPEG-2 video with PAL resolution (720 × 576 pixels) at 25–30 frames per second (fps).

Image distortion: Radial distortion has to be compensated for every image sequence based on recordings of a calibration pattern. A printout of a chessboard pattern is made available and captured before the inspection by the surgeon or clinical staff.

Illumination: The organ surface is illuminated by the light source, which is part of the endoscopy system available in the operating room (halogen or LED). Inhomogeneous illumination has to be compensated for in a preprocessing step within the algorithm pipeline.

Image quality: The algorithm has to be able to deal with changing image quality, due to poorly textured tissue or floating particles and fluids.

Scene shape: Since the shape of the bladder is assumed to be spherical, the algorithm has to be able to generate and display spherical panorama images.

Camera motion: The endoscope is manually navigated by the surgeon according to the clinical routine. Thus, the motion of the camera is not constrained and subject to six degrees of freedom: translation along all three spatial axes and rotation around all three spatial axes.

Hardware: The algorithm is developed and evaluated on standard PC hardware with Intel® Core™ i7-2700K CPU at 3.5 GHz, 8 GB RAM.

Computational speed: The algorithm has to run in real-time, i.e. the panorama image has to be updated at full video frame rate (≥ 25 fps). The algorithm may be run in parallel threads, some of which produce approximate preliminary results at 25 fps to enable the real-time look and feel for the user while others improve upon the accuracy with noticeable but acceptable delay. The average delay is required not to exceed one second.

Scalability: The algorithm has to scale in such a way, that the requirements regarding computation time can be met for a typical cystoscopic sequence. The duration of the available cystoscopy recordings, used within this thesis, range from 8 to 70 seconds. The set of images which has to be processed by the algorithm consequently consists of up to $70\,\mathrm{s} \cdot 25\frac{\mathrm{frames}}{\mathrm{s}} = 1750$ frames.

Panorama size: The resulting panorama images need to present part of the bladder which is large enough to enable the urologist to visually relate the location of clinical findings with known anatomical landmarks.

User interaction: No user interaction can be required during the execution of the stitching algorithm, except for starting and stopping the process.

Parametrization: The algorithm may be adjustable by a set of parameters. One set of parameters should be suitable for all input sequences. Tuning of parameters to different input data is undesirable.

1.5 Main Contributions

The aim of this thesis is the design, development, and evaluation of a real-time stitching approach for cystoscopic images. This includes all algorithmic modules involved: image preprocessing, registration, global alignment, optimization, and visualization. The goal of the thesis is accomplished if 1) the developed algorithms meet all requirements mentioned above, 2) the new contributions made in each field outperform the state of the art, and 3) the resulting panorama images are deemed suitable by a group of experts to improve endoscopic examinations and their documentation.

The main contributions of this thesis to the field of real-time endoscopic image stitching are as follows:

Large mosaic generation in real-time: As will be further discussed in Chap. 2, state-of-the-art approaches for endoscopic image stitching accomplish only one of the two goals: either large panoramas are generated in computationally expensive offline procedures, or real-time processing is achieved for rather small panorama images. The approach developed within this thesis is the first algorithm which is able to construct larger panorama images in real-time to create a field of view which is valuable for the surgeon in terms of improved navigation and documentation. An outline of the algorithm is presented in Chap. 3.

Fast polynomial filtering: A new filtering approach is developed, which extends the concept of integral images. Generally, integral images have a great potential to accelerate image filtering operations. So far, this calculation method has been restricted to box filters or combinations of box filters. Chap. 4 presents a generalization of the concept of integral images, which allows to also accelerate filtering operations with polynomial filters. In this thesis, polynomial filters are applied for fast image smoothing in the context of correcting inhomogeneous illumination as well as for fast keypoint detection and description within a feature-based image registration approach.

Shading correction: A new shading correction method is presented, specifically tailored to endoscopic applications. While established methods presume a gray-world and tend to produce color artifacts for images with a dominant coloring, this algorithm takes into account that endoscopic images tend to be dominated by a red coloring. The new method, presented in Chap. 5, reduces shading caused by inhomogeneous illumination without producing undesirable color artifacts. The method is based on fast polynomial filtering, as described above.

Motion model: Classical photo stitching algorithms usually assume the captured scene to resemble a plane (without significant differences in depth) or the camera to perform a mere rotation around its optical center (without any translation). Both cases allow to use a perspective transform to accurately model the image motion between successive images. Since the bladder has a spherical shape and the endoscope undergoes unconstrained motion, Chap. 6 considers different motion models with regard to their suitability for spherical image stitching. It will be shown that the perspective motion model produces systematic errors when applied to a spherical scene under arbitrary camera motion. Therefore, an alternative motion model is developed, which is better suited for spherical image stitching.

Feature-based registration: Since fast and robust image registration is a crucial part of any stitching algorithm, the development of a suitable registration method accounts for a major part of this thesis. After evaluating established feature detection and extraction methods regarding their suitability for endoscopic images in Sec. 7.2, a new feature detection and description algorithm is presented in Sec. 7.3. Key concepts of this method are the application of polynomial filtering to implement fast and robust feature detection and description, a hierarchical non maximum suppression scheme to accelerate the keypoint detection process, and a binary feature descriptor which allows to efficiently match keypoints between images.

Parallel tracking and stitching (PTAS): Parallel tracking and mapping (PTAM) is a well-established concept to implement algorithms for real-time camera localization and scene reconstruction. It separates the two main challenges camera tracking and scene mapping into disjoint processes which exchange information relevant to both tasks. This concept is adapted for real-time image stitching, disjoining the processes of feature tracking, image stitching, and global panorama optimization. The parallel tracking and stitching framework (PTAS) is presented in Chap. 8.

Chapter 2

State of the Art in Endoscopic Image Stitching

2.1 General Image Stitching

[1]Image stitching refers to the process of (automatically) creating panorama images by combining single digital images on a common canvas. The methodology of image stitching has its roots in the photography and photogrammetry community. In the 1980s and 1990s, an increasing number of scientists were exploring ways of automatically registering images obtained from a photo or video camera and generating globally consistent maps with a wide field of view. Early fields of application were the creation of high-resolution images from satellite photos and photo-mosaics from digital cameras. Fig. 2.1 shows a photo mosaic of the castle garden in Erlangen taken with a smartphone camera and stitched with the integrated photo app.

Image stitching algorithms are readily available in a number of commercial tools. Image editing suites like Adobe Photoshop, as well as dedicated stitching software like Microsoft Photosynth and AutoStitch provide off-the-shelf solutions to generate panorama images, primarily developed for photo collections captured with digital cameras [ZCT$^+$14]. AutoStitch has been developed by Brown and Lowe [BL07]. Microsoft Photosynth is based on work presented by Snavely et al. [SSS07]. Generally, the set of images to process may either be an unordered collection of photographs or stem from a video stream, which implies a temporal ordering and only little change between successive frames.

2.1.1 Image Stitching Pipeline

The majority of image stitching (or mosaicking) approaches share a common basic *pipeline* of procedures. The general concept for creating image mosaics has been described by Szeliski et al. [SS97]. Fig. 2.2 illustrates the individual steps involved in the mosaicking process.

If necessary, each image is *preprocessed* to facilitate further processing. For endoscopic applications, typical preprocessing steps are detection of the relevant

---

[1] An extended version of this chapter has been published in [BW16] © 2016 IEEE.

Figure 2.1: Panorama image of the castle garden in Erlangen stitched from multiple photographs.

image region (aperture mask), compensation for radial distortion, and reduction of vignetting effects (shading correction). Often, the amount of image data is further reduced by subsampling the input images at the beginning of the pipeline. The subject of shading correction is discussed in more detail in Chap. 5. Distortion correction is usually part of a camera calibration procedure, which is the topic of Sec. 3.2.

Image registration refers to the process of finding a transformation that matches images to each other in a pairwise manner. This is one of the most crucial steps in the mosaicking process. Registration algorithms can be categorized into *pixel-based* and *feature-based* approaches. Pixel-based algorithms try to minimize an optimization criterion calculated over the entire set of pixels within the overlap regions of the two images. Common criteria are the *sum of squared differences* (SSD), *normalized cross-correlation* (NCC), and *mutual information* (MI). Feature-based approaches extract higher-level features from the images, which are matched on the basis of their similarity. SIFT [Low04] and SURF [BETVG08] have proved to be among the most distinctive features and have been used in several stitching approaches [BL03, SSS07, ZCT$^+$14]. Further algorithms for feature-based image registration are discussed in Chap. 7. When streaming images from a video source, registering every single frame to its predecessor results in a heavy computational load as well as an unnecessarily high amount of overlap between frames. Therefore, most authors choose to implement some sort of frame selection mechanism, using either a fixed or an adaptive step size.

A motion model is needed to establish the mathematical relationship between pixel coordinates of overlapping image pairs. The one most commonly used is a perspective transformation (homography $\boldsymbol{H} \in \mathbb{R}^{3\times3}$) in the two-dimensional projective space $\mathbb{P}^2$. The homography accurately describes a coordinate transformation between two views: 1) if the scene is planar, or 2) if the camera motion between the views is a pure rotation (without translation). For a general perspective motion model, $\boldsymbol{H}$ is a transform with eight parameters. Other options for inter-frame motion models are simpler translational or affine models (with two and six parameters, respectively). In the case of unconstrained camera motion combined with non-planar scene geometry, more complex motion models are necessary to accurately describe the resulting image motion. A quadratic model with twelve parameters allows for quadric scene surfaces viewed by a weak perspective camera [HZ04, p. 171]. To model image motion between two perspective projections of a quadric scene, a non-

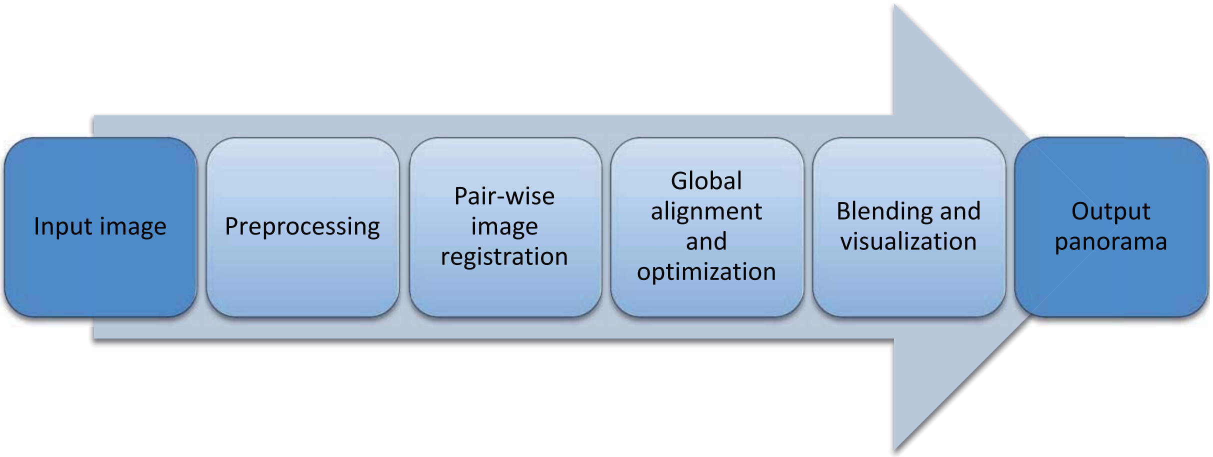


Figure 2.2: The common pipeline for image stitching represents the basis for many mosaicking and view expansion algorithms. Each selected input image undergoes a preprocessing step. Pairs of input frames are registered to enable the alignment of all frames in a common coordinate system. Blending between images reduces seam artifacts at the frame borders. The result of stitching several input frames is one panorama image of the scene. [BW16] © 2016 IEEE

linear transformation with 17 degrees of freedom is required [ST97]. More details about different motion models are given in Chap. 6.

In the case of pixel-based registration the image transform parameters are estimated by an optimization procedure. In the case of feature-based registration, the point correspondences give rise to an over-determined system of equations, which can be solved using a Random Sample Consensus (RANSAC) scheme [FB81] or one of its numerous derivatives, such as MSAC or MLESAC [TZ00] (ref. Sec. 7.1.9). For comprehensive surveys on general aspects of image registration the reader is referred to the publications by Brown [Bro92], Goshtaby [Gos05], and Zitová and Flusser [ZF03].

The pairwise-registered images are then aligned in a common *global* coordinate system by projecting the images onto a surface, which is usually planar, cylindrical, or spherical. For linear motion models, images can be stitched incrementally by concatenating the individual inter-frame transformations. However, this will inevitably lead to an accumulating error so that it only allows to stitch rather small patches. For larger mosaics, global consistency can be achieved by reducing a global error measure. This implies solving a global *optimization* problem. The most common approach is bundle adjustment (BA), presented by Triggs et al. [TMHF00]. BA was originally developed for 3D reconstruction using structure-from-motion algorithms. It was further adapted for image stitching by Shum and Szeliski [SS01]. BA is inspired by the idea that the set of projections of scene points onto camera images can be interpreted as a bundle of rays. During bundle adjustment, the scene point coordinates and camera parameters are refined simultaneously to optimize the global projection error. The error is usually defined in terms of the deviation of image coordinates of projected scene points from their observations. The subject of global optimization is further discussed in Sec. 8.4.