Three-Dimensional Surface Reconstruction of the Bladder from Cystoscopic Video: An Initial Investigation

Timothy D. Soper[a], Michael P. Porter[b], Eric J. Seibel[a]
[a]*Department of Mechanical Engineering, University of Washington, Seattle, WA*
[b]*Department of Urology, University of Washington, Seattle, WA*

ABSTRACT

Flexible cystoscopy is the most common procedure performed by urologists, and thus imposes considerable time and financial cost. In this work, we present custom software capable stitching full 3D mosaics of the bladder for more expedient examination. The software is part of an automated bladder scanning system that includes a miniature (1.5mm OD) flexible endoscope and a computer-controlled steering mechanism. The intent of this system is to relieve the urologist of performing surveillance cystoscopies while opening avenues for remote interpretation and longitudinal assessment. Unlike other approaches that are limited to 2D image alignment strategies, our stitching algorithm computes a 3D stitched surface model of the bladder from cystoscopic video using bundle adjustment. The computer model can then be reviewed by the urologist post-procedurally at a later time or remote location. Our software was tested on cystoscopic video of an excised pig bladder. The resulting reconstruction possessed a projection error of 1.66 pixels on average and covered 99.6% of the bladder surface area. The software was further tested on high definition video from rigid cystoscopy in a patient. Though the software was unable to reconstruct a 3D model from, a planar panorama of the posterior bladder wall was achieved with an average projection error of 4.45 pixels. Results from this initial investigation demonstrate the feasibility of using the 3D image stitching software for virtual bladder examination as part of an automated bladder scanning system.

Keywords: bladder, endoscopy, cystoscopy, structure from motion, bundle adjustment, image stitching

1. Introduction

Flexible cystoscopy is the most common procedure performed by urologists in the United States, most commonly as part of routine surveillance for bladder cancer. This is due to both the associated prevalence (5th among all cancers), and a 50% recurrence rate (Jemal et al., 2010). During the examination, the urologist advances the flexible cystoscope through the patient's urethra into the bladder, manually steering the distal tip to inspect the lining of the bladder. Given the limited field of view, urologists must methodically maneuver the scope over each region to ensure a complete examination of the bladder wall.

Though flexible cystoscopy is the current gold standard for routine bladder surveillance, there are drawbacks for the both the patient and physician. First, conventional cystoscopes can impose moderate to severe discomfort in patients who must accommodate scope diameters of 5mm or larger under local anesthesia. Though smaller endoscopes are desirable, miniaturization requires reducing the number of pixel elements (optical fibers or sensor elements), thereby degrading image resolution and field of view, and making cancer detection more challenging. Second, the urologist must carefully examine the entire bladder under limited visualization, inspecting for early signs of recurrence. If a suspect lesion is identified, the patient must then be rescheduled for biopsy or resection of the mass during a second

cystoscopy performed under general anesthesia. These procedures can constitute a significant percentage of the urologist's time and clinical resources.

Our laboratory has developed a novel ultrathin scanning fiber endoscope (SFE) that may potentially transform the current clinical standard in bladder surveillance (**Figure 1**). Unlike conventional CCD- or fiber bundle-based endoscopes, the SFE acquires high resolution images via piezo-actuation of a single laser scanning optical fiber housed within a 1.5mm distal tip (Lee et al., 2010; Seibel et al., 2006). Because of its small size, the SFE could be more easily tolerated. Furthermore, the small size and flexibility make the SFE highly amenable to automation, where it could be mechanically articulated to scan the entire inner surface of a bladder (Yoon et al., 2009). Together, these technological components could relieve the urologist of manually performing each and every surveillance procedure. In this new approach, surveillance cystoscopies could be conducted with limited oversight by a nurse or ancillary care provider, allowing the urologist to analyze endoscopic image data post-procedurally, thereby streamlining clinical workflow. Not only would this promote more efficient interpretation, but it would also present opportunities for remote medicine and longitudinal assessment.

Here, we present a 3D image stitching algorithm that is aimed at constructing digital surface models of the bladder for expedient review urologists. Though endoscopes provide only a limited field of view of the bladder, the software stitches the set of images together to generate a full 360° panorama of the bladder. Such a panorama would provide full-field visualization of the bladder, creating greater anatomical context within which to detect and evaluate disease while compressing lengthy video segments into a single 3D stitched surface model. Whereas most panoramic stitching programs can only be applied to limited regions of the bladder, the software presented here utilizes *structure from motion* to fully reconstruct the surface of the bladder. Individual video frames are then stitched onto this surface, yielding a texture-mapped surface model of the bladder that can be examined in 3D. It is believed that a 3D bladder panorama will greatly expedite expert examination and provide visual assurance that no region was missed. Though this is proposed as part of a new clinical bladder surveillance system, 3D mosaics of the bladder may be reconstructed from conventional cystoscopy for maintaining digital records of each examination. In the future, such techniques may prove valuable in endoscopy of other organs.

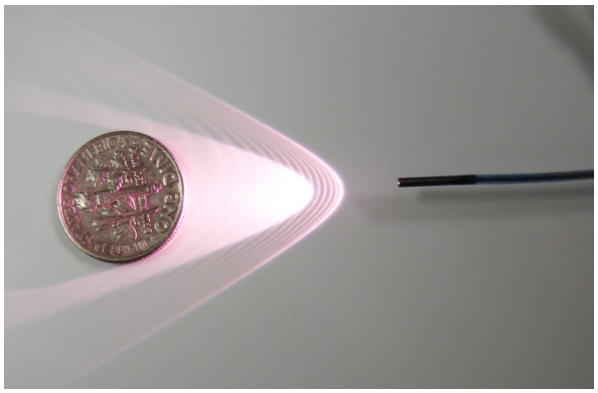


Figure 1: SFE imaging of a dime. The rigid distal tip of the highly flexible shaft is 1.5mm OD and 9mm in length.

2. Related Work

Panoramic stitching or *photomosaicking* is the process of combining overlapping images onto a common compositing surface to construct expanded views. Clinically, stitching has been used to merge multiple fluoroscopic x-ray images (Yaniv and Joskowicz, 2004), breast ultrasound images (Chang et al., 2010), and endoscopic images of the retina (Seshamani et al., 2006), esophagus (Seibel et al., 2008), abdomen (Naya et al., 2009), and bladder (Behrens, 2008).

2.1. Cystoscopic Image Stitching

Mosaicking of cystoscopic images provide urologists wide field visualization of the bladder. These panoramic views serve as valuable frames of reference in which to navigate the bladder (Behrens et al., 2011), determine the spatial distribution of multifocal tumors (Olijnyk et al., 2007), and to evaluate changes during follow-up examination. The synthesis of wide field views is also highly beneficial in fluorescence imaging of the bladder, which requires the urologist to maintain a short working distance to acquire sufficient fluorescence signal (Behrens, 2008). However, current approaches to stitching cystoscopic images are limited in that they do not extend to full 360° views covering the entire bladder, as is desired in our proposed application. This is primarily for two reasons: first, most panoramic stitching applications employ only 2D-2D image alignment methods; and second, stitching is commonly performed by local rather than global alignment.

2.1.1. 2D Alignment

The alignment of any pair of images requires some known mapping between a pixel coordinate x in one image to a coordinate x' in the other. In photomosaicking, such a mapping is frequently calculated on the basis of a presumed model of the camera motion and scene structure. In the construction of flat panoramas, the scene of interest is assumed to be planar such that images are related by a common homography. Alternatively, spherical or cylindrical panoramas can be generated from images in which the camera undergoes pure rotation, where landmarks lie on a plane at infinity.

In the case of images acquired from a freely moving endoscope within a nonplanar bladder, such models are inadequate. However, the bladder surface can be modeled as piecewise planar over a limited field of view. **Figure 2a** depicts a flat panorama of the bladder constructed from a sequence of video frames successively aligned for more wide field visualization. Over larger fields of view, the curvature of the bladder begins to stretch images as they are added, resulting in a severely distorted map of the bladder wall (**Figure 2b**). Furthermore, because the bladder surface is not truly planar, some degree of misalignment is introduced between overlapping images. To compensate, multiple local panoramas may be constructed from video sequences covering individual segments of the bladder (Behrens et al., 2009), or by requiring a predetermined endoscope trajectory (Miranda-Luna et al., 2008). However, while these methods can be used for panoramic visualization, their implementation requires introduces additional layers of control that must be moderated by the urologist.

2.1.2. Local vs. Global Alignment

The panoramas of **Figure 2a** and **Figure 2b** were constructed using *local alignment* strategies, where for each image I_k, we compute the local transformation $T_{k,k-1}$ that aligns I_k with the preceding image I_{k-1}. Thus, the global transformation T_k that aligns I_k with the panorama and I_0 can be easily computed as a chain of local transformations where:

$$T_k = T_{k,k-1} T_{k-1} = \prod_{i=1}^{i=k} T_{i,i-1} \quad (1)$$

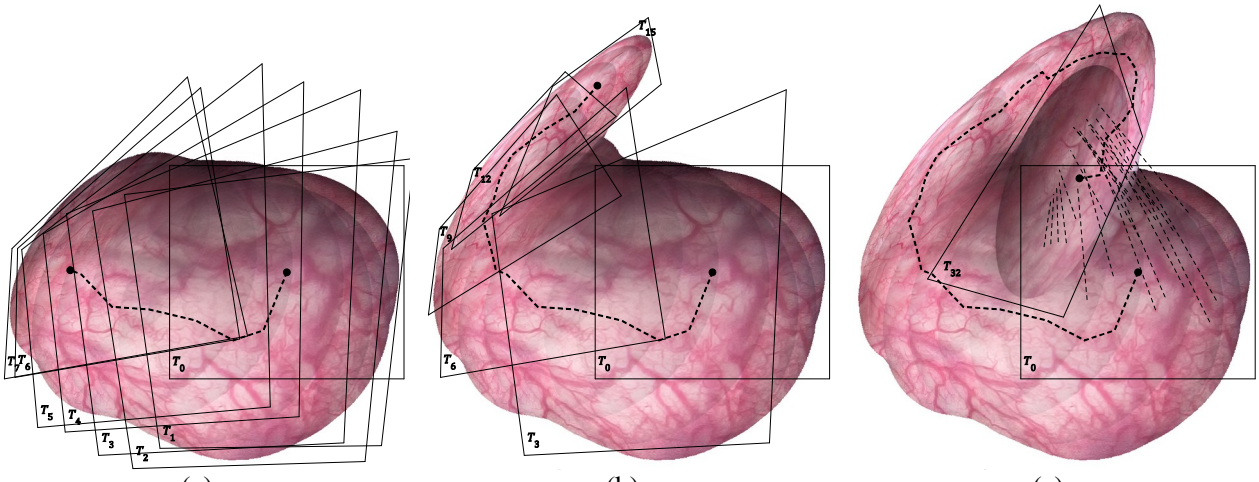


Figure 2: Panoramic stitching of cystoscopic video sequences by local alignment. Seven images are globally aligned and stitched by computing the local transformation chains $T_0 - T_7$ (a) with a dashed line connecting frame centers. Over a longer sequence, added frames begin to distort and stretch due to the non-planar bladder surface shape as seen by the transformation T_{15}(b). When the endoscope motion forms a loop, overlapping frames at the start and end of the loop are misaligned as a result of propagating local alignment errors (c). The misalignment of the transformation T_{32} is depicted by the straight dashed lines that connect matching features.

Though local alignment produces consistent panoramas over sequences strips such as those in **Figure 2a & b**, it is not robust to loops in the endoscope motion path. **Figure 2c** illustrates the misalignment that results from a loop in the motion path between a frame I_k and the first frame I_0. This is the result of small alignment errors that propagate through the transformation chain of (1). This particular issue, known in photogrammetry as *gap closure*, necessitates some form of global alignment, in which the entire set of transformation parameters contained in T_0 through T_k are computed simultaneously. This requires that the set of transformations T be optimized over all overlapping frame pairs and not just between subsequent frames. Thus significant processing is necessary to both detect all overlapping frame pairs and optimize, by some iterative means, the entire set of transformations. In applications that require realtime stitching, global alignment is not feasible. Miranda Luna *et al*. proposed a modified local alignment strategy that searches for loops in the endoscope motion path, and once found, redistributes alignment error over the entire set of frames (Miranda-Luna et al., 2008). However, to cover the entire bladder, multiple sweeps of the bladder are necessary, resulting in a large patchwork of overlapping images that must be simultaneously registered by means of global alignment.

2.2. Structure from Motion

The 2D-alignment methods discussed previously are generally limited to planar compositing surfaces. To generate a full 3D stitched surface model, the scene geometry must be simultaneously recovered. Using tomographic image data, such as computed tomography (CT) or magnetic resonance images, endoscopic images can be fused with extracted surfaces of (Dey et al., 2000; Rai and Higgins, 2009).

However, in this application, acquisition of a preoperative CT scan is not assumed, nor is this strategy to be ideal for the bladder, which lacks distinguishing surface features and is highly distensible. Alternatively, given a moving video sequence, scene geometry can be recovered from the image data directly. This process, termed structure from motion (SfM), simultaneously reconstructs 3D feature points along (structure) and camera poses (motion) from a set of 2D images with known correspondences.

2.2.1. Feature Alignment

Reconstruction of scene and motion parameters is driven by the alignment of known image correspondences. The pixel position x_{ij} is determined by the 3D position of a feature point p_i, the pose of the camera described by a rotation and translation R_j and t_j, and the intrinsic calibration parameters K, yielding a generic function:

$$x_{ij} = f(p_i, R_j, t_j, K)$$

The position of a point p_i in world coordinates is first converted into relative camera coordinates by a rigid 3D transformation, given by:

$$p'_{ij} = R_j p_i + t_j \tag{2}$$

The normalized pixel coordinates ($\hat{x}, \hat{y}$¿ projected onto the camera at the plane $z=1$, is calculated as $\hat{x}_{ij} = p'_{ij}(x)/p'_{ij}(z)$ and $\hat{y}_{ij} = p'_{ij}(y)/p'_{ij}(z)$. In most instances, the lens of the camera introduces some degree of radial distortion, commonly modeled by the nonlinear quadratic:

$$\begin{aligned} \hat{x}' &= \hat{x}(1+\kappa_1 r^2 + \kappa_2 r^4) \\ \hat{y}' &= \hat{y}(1+\kappa_1 r^2 + \kappa_2 r^4) \end{aligned} \tag{3}$$

where $r = \sqrt{\hat{x}^2 + \hat{y}^2}$. The normalized pixel position is finally converted to the pixel position (x, y¿ in image coordinates by the equation:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \begin{bmatrix} \hat{x}' \\ \hat{y}' \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \alpha f_x & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}' \\ \hat{y}' \\ 1 \end{bmatrix} \tag{4}$$

where K is the camera calibration matrix constructed from the focal length (f_x, f_y), skew coefficient α, and center pixel (x_0, y_0).

2.2.2. Bundle Adjustment

In the case of the two-view geometry, SfM can be computed directly from linear methods, such as the normalized 8-point algorithm of (Hartley, 1997). For *N*-view geometries, where *N* may represent several hundreds or thousands of video frames, such large-scale SfM is resolved through an iterative optimization

process referred to as bundle adjustment. Using this technique, databases of thousands of photographs have been used to reconstruct major landmarks for virtual tourism (Snavely et al., 2006). More recently, bundle adjustment has been applied to endoscopic image data to achieve super-resolution panoramas (Hu et al., 2010a) and in the reconstruction of organ surfaces (Hu et al., 2010b). For a more thorough discourse of bundle adjustment theory and its applications, we recommend reviews of the subject by (Triggs et al., 1999) and (Engels et al., 2006).

The solution produced by bundle adjustment is that which minimizes the measurement error or *pixel projection error* between the observed pixel positions x and predicted pixel positions $\tilde{x}$ computed from (2)-(4). **Figure 3a** illustrates the point-camera correspondences that arises from multiple observations of features. Defining the pixel projection error $z_{ij}=x_{ij}-\tilde{x}_{ij}$, where z is a vector containing the error of every observed point-camera pair, a cost c can be computed as the sum of squared errors.

$$c=\frac{1}{2}z^T z \tag{5}$$

This cost can be defined as a function $c(x)$, where x is a state vector containing the entire concatenated list of point and camera parameters, given by $x=\left[p_1^T,\ldots,p_M^T,\omega_1^T,t_1^T\ldots,\omega_N^T,t_N^T,f_x,f_y,u_0,v_0,\kappa_1,\kappa_2\right]^T$. Here, the rotation matrix R describing the camera orientation is replaced with an orientation vector ω computed from the Rodrigues rotation formula. Locally modeling the cost function by the quadratic Taylor expansion,

$$c \text{¿} \tag{6}$$

the update Δx that minimizes c ¿ is given by the nonlinear least-squares solution:

$$H\Delta x=-g \tag{7}$$

where g is the gradient and H is the Hessian matrix of $c(x)$. Both can be expressed in terms of the Jacobian J of the measurement error z, where

$$J=\begin{bmatrix}\frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_1}{\partial x_n}\\ \vdots & \vdots & \vdots\\ \frac{\partial z_m}{\partial x_1} & \cdots & \frac{\partial z_m}{\partial x_n}\end{bmatrix},\ g=Jz,\ H\approx J^T J$$

By solving for Δx from Error: Reference source not found, the structure and motion parameters are refined iteratively, where $x^{i+1}=x^i+\Delta x$, recomputing H and g at each update. Considering the state vector x can contain several thousand parameters, solving for Δx is not straight-forward. However, the system of equations in (7) can be broken into its point and camera based constituents, yielding:

$$J=\begin{bmatrix}J_p & J_c\end{bmatrix}, H=\begin{bmatrix}H_{pp} & H_{pc}\\ H_{pc}^T & H_{cc}\end{bmatrix}=\begin{bmatrix}J_p^T J_p & J_p^T J_c\\ J_c^T J_p & J_c^T J_c\end{bmatrix}$$

producing the sparse banded matrix seen in **Figure 3b&c**. Conveniently, the diagonal structures of H_{pp} and H_{cc} are easily inverted, allowing reformulation of (7) to a reduced camera system:

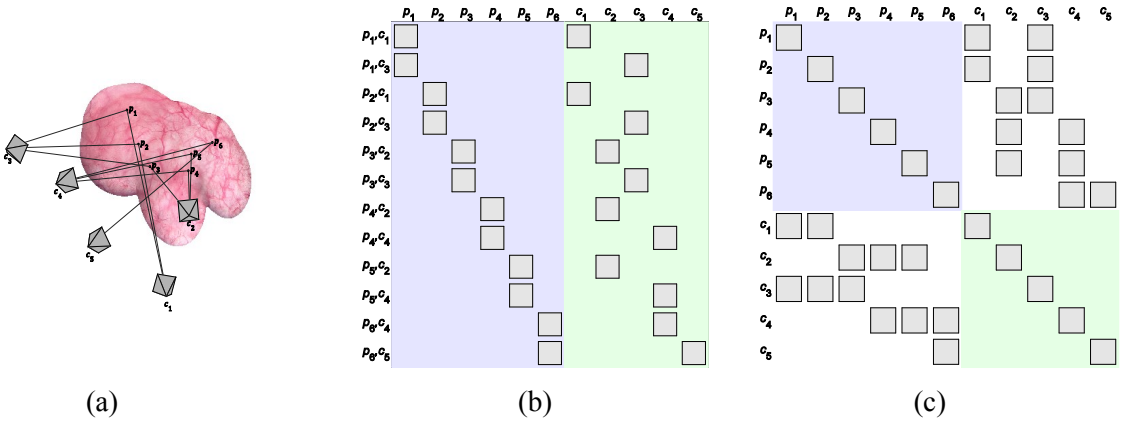
$$\overline{H}_{cc}\Delta x_c=-\overline{g}_c \tag{8}$$

where $\overline{H}_{cc}$ is the *Schur Complement*, computed as $\overline{H}_{cc}=H_{cc}-H_{pc}^T H_{pp}^{-1} H_{pc}$, and $\overline{g}_c=H_{pc}^T H_{pp}^{-1} g_p-g_c$. The update to the camera parameters Δx_c is then used to compute the update to the point parameters:

$$\Delta x_p=H_{pp}^{-1} g_p-H_{pp}^{-1} H_{pc}\Delta x_c \tag{9}$$

The solutions produced by (8) and (9) are unique to bundle adjustment problems in that very large systems of equations, potentially containing thousands of points and cameras, can be solved by leveraging sparsity introduced by two sources. First, structure from motion is a bipartite problem in which structural and camera parameters do not combine. Put another way, each measurement projection error z_{ij} is the result of only one point $i$and one camera j. Second, in most cases, each camera sees only a small subset of the total number of feature points. Lastly, to ensure that each adjustment produces an improved estimate of x, step control is implemented in the form of the Levenberg-Marquardt algorithm, where (7) is rewritten to include a variable weighting factor λ:

$$\left(H+\lambda\, diag(H)\right)\Delta x=-g \tag{10}$$



(a) (b) (c)

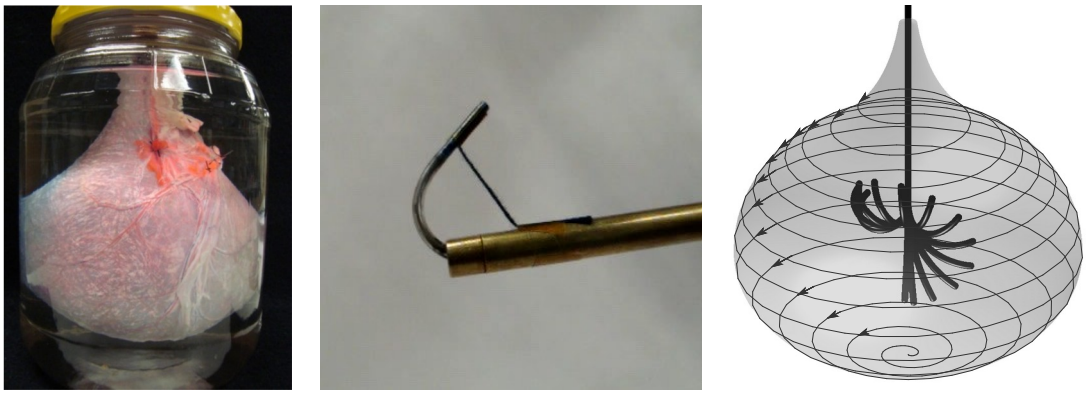
Figure 3: An example of camera and point correspondences are illustrated over a small surface patch in the bladder (a). The resulting correspondences produce a sparse Jacobian matrix (b) containing the

derivatives $\partial z/\partial x$, where each entry denotes the track-camera pair p_i, c_j. The Hessian matrix $H = J^T J$ (c) contains the Gauss-Newton approximations to the second derivatives of the cost function c. The strictly point-based and camera-based matrix entries are signified by the blue and green regions, respectively.

3. 3D Surface Stitching of Cystoscopic Video

3.1. Cystoscopy of an Excised Pig Bladder

SFE imaging was conducted in an excised pig bladder (**Figure 4a**). Following resection, dye was injected into the two main arteries that feed the two hemispheres of the bladder to maintain vessel contrast during cystoscopy. Red dye was injected into one artery and blue dye into the other. The bladder was then fixed in formalin and suspended within a jar. An access port was inserted into the jar lid and attached to the bladder to allow for direct introduction of the SFE. To achieve a full 180° bend angle, the SFE was tethered to a rigid introducer tube (**Figure 4b**). The jar and bladder were then placed on a rotating stage. Images were acquired in a spiral path as the SFE was slowly extended to bend back toward the bladder neck (**Figure 4c**). A frame grabber was used to save cystoscopic images to a Dell 470 Precision Workstation (3.40 GHz CPU, 4GBytes RAM) for further processing. Because video was taken from inside a dyed pig bladder using our SFE, acquired images (depicted in subsequent sections) do not closely resemble those from conventional cystoscopy in a patient.



(a) (b) (c)

Figure 4: A pig bladder was excised and fixed in formalin (a). To maintain vessel contrast, arteries were separately injected with red and blue ink. Cystoscopy of the bladder was performed using the SFE. To achieve 180° back bending, the SFE was tethered to a rigid insertion tube (b). A spiral scan path (c) was used to image the entire inner surface of the bladder.

3.2. 3D Surface Stitching Software

Figure 5 outlines the processing steps of the custom panoramic stitching software. The code was completely developed in MATLAB for the reason that it provides support for sparse matrices, which is critical to bundle adjustment. Moreover, it contains functions for reading video files and image processing. The pipeline is broken into two stages: frame analysis and 3D reconstruction. For each image features are detected and characterized for each video frame as the basis for pairwise frame matching. From the total number of frames, a sparse subset of frames are selected to eliminate redundancy and

economize on processing time. From this collection of frames, a sparse global match search is conducted to identify all overlapping frame pairs. Image features that are consistently matched between all frames are identified as 3D *tracks*. From the set camera frames and surface points, structure from motion is resolved incrementally through non-linear least squares using bundle adjustment. From this reconstruction, we compute an explicit model of the bladder surface and finally stitch the individual images onto the surface and blending overlapping regions. Each stage of the 3D stitching software is described in conjunction with results from processing of the SFE cystoscopy of the excised pig bladder.

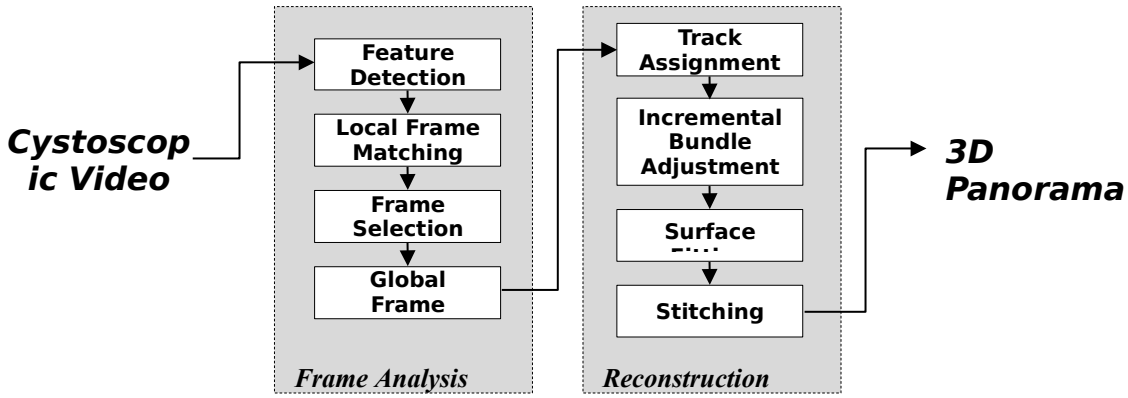
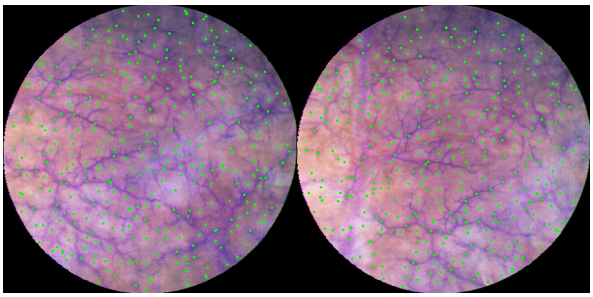


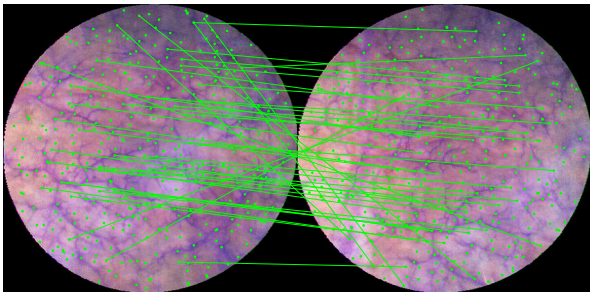
Figure 5: The 3D surface stitching software pipeline

3.3. Feature Detection/Description

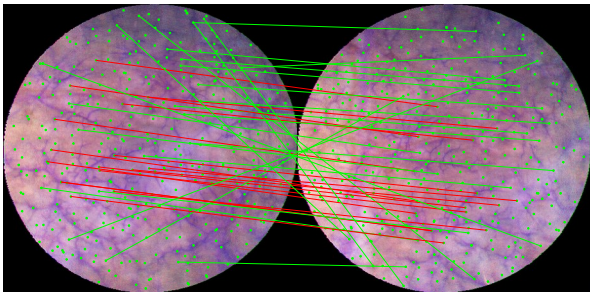
The endoscopic video consists of several hundreds or thousands of frames that overlap as part of a comprehensive and consistent image patchwork covering the entire inner surface of the bladder. To accurately construct a full stitched surface model, overlapping frames must be matched. These frame matches are established through the correspondence of a discrete set of features. For each video frame, a set of scale- and rotationally-invariant features are detected using the well-known scale-invariant feature transform (SIFT) (Lowe, 2004). This is implemented using the VLFeat toolbox (www.vlfeat.org) available for MATLAB, whose modified SIFT detection function takes an image as input and returns a set of features as in **Figure 6a**. Because SIFT features are computed from grayscale images, only the green image channel is used as input from the SFE video frames as appears to contain the best contrast for blood vessels. Each feature is characterized by a *keypoint* $x=(x,y)$ in pixel space and by a 128-element description vector D. Conveniently, these features combine to help identify corresponding image regions, and constitute discrete 3D feature points from which the bladder surface shape is reconstructed.



(a)



(b)



(c)

Figure 6: Alignment of two overlapping video frames acquired from the SFE inside an excised pig bladder. SIFT keypoint features are first detected in both images and characterized by a description vector D (a). Features are then matched on the basis of the separation distance in feature space (b) as shown by the connecting green lines. Matches are finally pruned by RANSAC leaving only those feature matches that are related by a common homography as shown by red lines (c).

3.4. Local Frame Matching

To begin the frame matching process, each frame is first matched to the subsequent frame. Though local alignment of sequential images can be computed from direct comparison of pixel intensities (Hernandez-Mier et al., 2010) or mutual information (Miranda-Luna et al., 2008), such methods are more computationally exhaustive. Alternatively, feature-based frame matching is achieved in two steps: feature matching and pruning by *RANdom SAmple Concensus* (RANSAC). In attempting to match a pair of frames $I \leftrightarrow I'$, the features themselves are first compared by computing the Euclidean distances $\left\|D_i - D_j^{'}\right\|$ for each possible feature pair (i, j). A pair (i, j) is determined to match if both are mutual nearest neighbors separated by a distance $d_{nn} = \left\|D_i - D_j^{'}\right\|$ in feature space, and if

$$\frac{d_{2nn}}{d_{nn}} > \epsilon$$

where d_{2nn} is the second nearest neighbor distance of either feature and ϵ is a predetermined threshold greater than 1 (**Figure 6b**).

For local matching of sequential frames, rigorous calculation of every possible feature pair is not necessary. Rather, each feature is compared only to those features falling within a limited subregion of the two images. This more expedient approach, known as *guided matching*, leverages the fact that the endoscope undergoes only limited motion between subsequent frames, and thus displacement of features

can be presumed to be small. Here, we impose a maximum displacement equal to 25% of the frame width for local matching of sequential frames.

In the second step, a geometric transformation function T is computed from the spatial arrangement of the matched features such that $x' = T(x)$. Because not all matches are consistent with a single transformation, T is computed on the basis of the greatest number of inliers arising from RANSAC (**Figure 6c**). For this application, the transformation relating matched feature keypoints $x \leftrightarrow x'$ is modeled as a planar homography H, such that

$$\begin{bmatrix} x'\omega \\ y'\omega \\ \omega \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Finally, the total number of feature matches or inliers determine the validity of the frame match $I \leftrightarrow I'$. Here, we use a minimum of 16 feature matches to establish frame overlap.

3.5. Frame Selection

To minimize processing time, only a subset of n images I_0 to I_{n-1} are selected from the total number of video frames I^v. For greater expediency, selection is performed concurrently with the local matching algorithm of 3.4. For each video frame I_k^v selected, the image is added at I_n, where n is the number of frames selected thus far, and k is added to a vector k that holds the video frame indexes of each frame selected. The frame selection process then continues through the following steps:

Step 1: Initialize selected frames $I_0 = I_0^v, I_1 = I_1^v$ and index vector $k = [0\ 1]$. Also initialize $n = 2$, $k = 2$, and $\Delta k = 1$.

Step 2: Read the new frame I_k^v.

Step 3: Attempt match with second to last selected frame $I_{n-2} \leftrightarrow I_k^v$. If a match is found, replace the last selected frame by setting $I_{n-1} = I_k^v$ and $k_{n-1} = k$, then **GOTO 7**.

Step 4: Attempt match with last selected frame $I_{n-1} \leftrightarrow I_k^v$. If a match is found, add $I_n = I_k^v$ and $k_n = k$, increment n and **GOTO 7**.

Step 5: No match was found. If $\Delta k = 1$, this is already the smallest interval, so continue with the next frame. **GOTO 8**.

Step 6: Halve the interval by stepping back, $k = k - \Delta k$ and then halve Δk and **GOTO 8**.

Step 7: Update the frame interval so that $\Delta k = k_{n-1} - k_{n-2}$

Step 8: $k = k + \Delta k$

Step 9: If $k \leq$ the total number of video frames, **GOTO 2**

3.6. Global Frame Matching

To suitably stitch a full panorama of the bladder, it is necessary to perform some global alignment of the entire collection of frames by way of a common set of visible features. Global alignment at this level requires frame matching over the entire set of images. Considering the total number of selected frames n

is commonly on the order of several hundred or thousand, the total number of possible frame matches is calculated as $(n¿¿2-n)/2¿$. To avoid an exhaustive global search of every possible frame pair, we institute a more economical approach that uses a sparse match search in three stages: sequential, nonsequential, and associated frame matching.

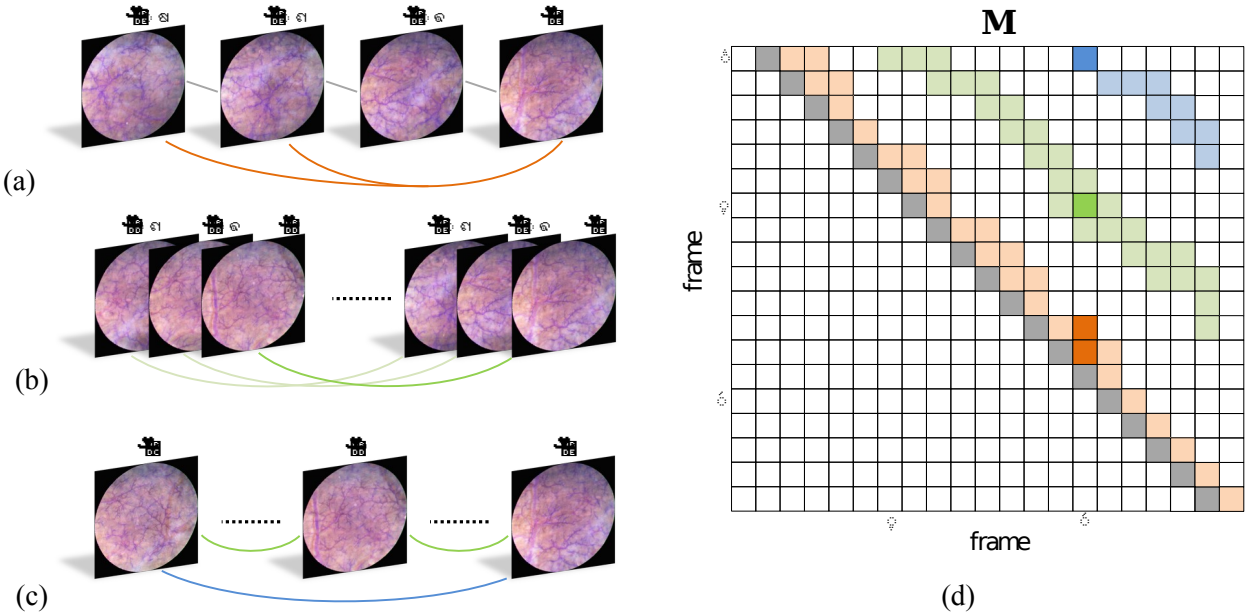


Figure 7: The three global frame matching stages are illustrated in examples (a-c) and on the match table M(d). First each frame is sequentially matched to as many preceding frames as possible (a). Following, the software searches nonsequential matches of frames that are not sequentially linked and are separated by some minimum number of frames. For each nonsequential match, recursive neighbor matches are sought, exposing large seams of frame overlap as illustrated by the first off-diagonal in green (d). An associated match search identifies matches between frames that mutually overlap a third frame (c). This uncovers additional seams of overlap as shown by the second off-diagonal in blue (d).

Following frame selection, the software possesses a set of several hundred or more video frames, each one locally matched to the subsequent frame. A sparse $n \times n$ match table M is first constructed to store the match data between any such pair of frames (**Figure 7d)**. In the first step, each frame is matched sequentially to as many preceding frames as possible until a match fails (**Figure 7a**). This creates the main diagonal of the match table in M. In the second step, the software searches for nonsequential matches $I_j \leftrightarrow I_k$, defined as frames that overlap as a result of loops in the motion path of the endoscope (**Figure 7b**). Programatically, these matches are determined by frames which are not matched sequentially and are separated by some minimum number of frames. We use a minimum separation of 10 frames to define a nonsequential match. Such a match can also serve as a seed around which additional frame matches can be located. For a nonsequential match $I_j \leftrightarrow I_k$, the software additionally tests for potential neighbor matches $I_{j-1} \leftrightarrow I_k, I_{j+1} \leftrightarrow I_k, I_j \leftrightarrow I_{k-1}, I_j \leftrightarrow I_{k+1}$. Additional neighbor matches are recursively tested, eventually uncovering large seams of overlap between "sweeps" of the endoscope represented by the first off-diagonal section in the match table M. This recursive neighbor search is

further expedited using guided matching techniques. Once a seed match $I_j \leftrightarrow I_k$ is identified, it is possible to pre-estimate the alignment of a candidate neighbor match ($I_{j-1} \leftrightarrow I_k$ for example) given by the homography $H_{j-1,k} = H_{j,k} H_{j-1,j}$, where $H_{j,k}$ and $H_{j-1,j}$ are already known. From this pre-estimate of the alignment, a more economical subsearch of matching features can be conducted on the basis of keypoint location.

In the final step of the sparse match search, termed associated frame matching, the software further attempts to match any pair of frames associated by mutual overlap with a third frame (**Figure 7c**). Thus for any match $I_j \leftrightarrow I_k$, each frame I_i for which there is also a match $I_i \leftrightarrow I_j$, is considered a possible match $I_i \leftrightarrow I_k$ by association. This process further exposes regions of frame overlap in M.

3.7. Track Assignment

To this point, the processing stages have involved characterization of individual frames and subsequent pairwise matching. To initiate a global reconstruction of scene and camera geometries, a set of features consistently matched over multiple frames must first be extracted. These features, referred to as *tracks* provide the basis of the 3D surface geometry around which the bladder is ultimately shaped. For each frame I_k, each feature x_{ik} is analyzed for possible feature matches in any previous frame. If this feature is not matched in a previous frame, it is assigned to a new track point $x_{ik} \mapsto p_m$, where m is the label index of the last and newly added track point p. If x_{ik} is matched to a feature x_{hj} in a previous frame I_j then x_{ik} is assigned to the track p_l where $x_{hj} \mapsto p_l$. In the event x_{ik} matches to several features assigned to different tracks, those tracks are deleted and the associated features are disregarded. In this implementation, any track viewed less than three frames were removed from the reconstruction.

3.8. Incremental Bundle Adjustment

As with most large-scale optimization problems, bundle adjustment progresses toward a local minima, and thus requires proper initialization for accurate scene reconstruction. Given the number of variables, bundle adjustment is most commonly approached in an incremental fashion, beginning with only a small set of cameras and tracks to initialize the reconstruction. As structure and camera parameters are resolved, additional cameras and tracks are slowly added, effectively building up a model of the scene and camera geometry over time. For many applications, the bundle adjuster is initialized using two matched frames separated by a sufficiently large baseline distance from which to derive the rigid 3D transformation (up to a scale factor) that relates them (Snavely et al., 2006).

In our approach, the reconstruction leverages the fact that the bladder is continuous and enclosed, and presumed to lack any hidden surfaces. Thus, for simplicity, the bladder is approximated as a sphere in the initial reconstruction stage as described in previous work (Soper et al., 2011). Thus each track point is constrained to lie on a sphere with a unit radius. Similarly, each camera position is restricted to lie with the spherical surface. This constraint is easily implemented by reparameterizing the Jacobian in terms of the 2D spherical coordinates (θ, ϕ). The benefit of such a constraint is that it helps to stabilize the reconstruction until it is fully formed. If the geometry is left unconstrained, tracks and cameras on the could warp and flex around gaps in the model, likely allowing the optimization to deteriorate to a false local minima.

Error: Reference source not found illustrates the incremental bundle adjustment over multiple passes. The process begins with a single active camera, arbitrarily placed in the center of the sphere $(0,0,0)$ pointing in the z direction. Following, tracks and cameras are added to the reconstruction in

alternating fashion. In odd passes, tracks that are visible to any of the active cameras are added. The position p of any track is estimated by projecting the associated feature pixel of each camera in which they are viewed to the surface of the sphere and averaging. In even passes, cameras which contain five or more active tracks are activated. Their initial pose is estimated using the 5-point algorithm of Quan & Lan (Quan and Lan, 1999). After activating addition tracks or cameras, bundle adjustment of the subset of track and camera parameters is iteratively conducted until fully minimizing the cost of (5). In this way, additional cameras and tracks are infused in a hand-over-hand method, reoptimizing at each step, until

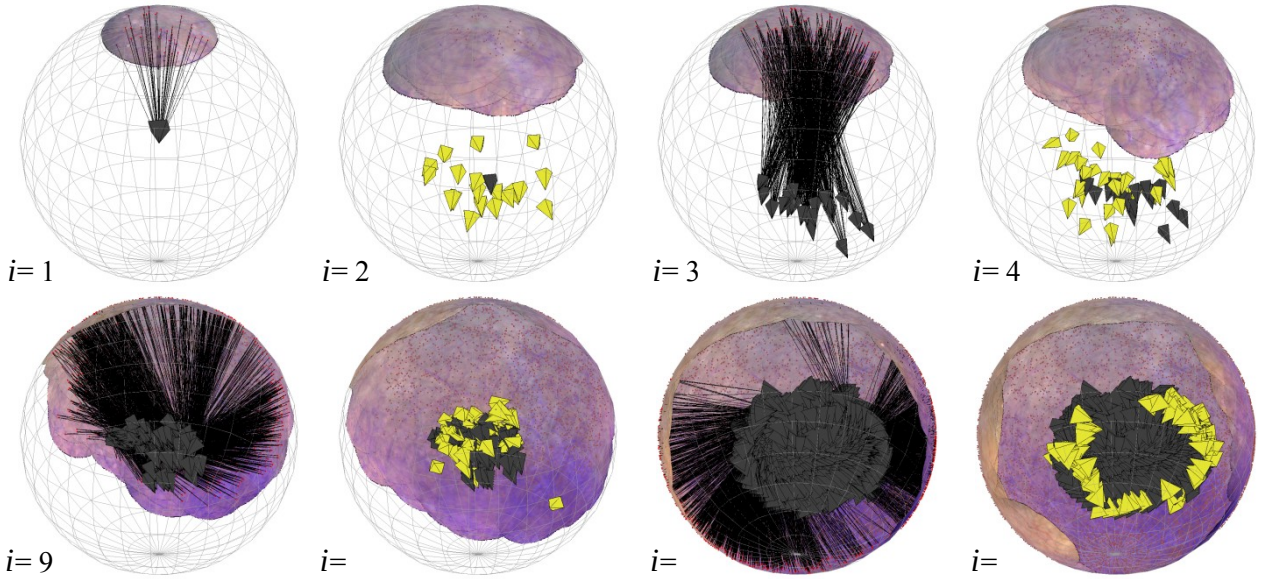


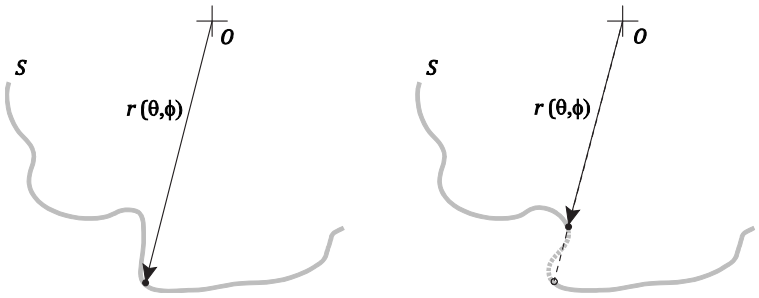
Figure 8: Several passes of the spherically constrained incremental bundle adjustment are depicted. The reconstruction is initialized to a single camera assumed to lie at the sphere origin and the set of tracks visible to this camera. In even passes, new cameras which view 5 or more of the active tracks are added to the reconstruction (shown in yellow). In odd passes, tracks visible to active cameras are added (red). In the beginning, the reconstruction appears nearly planar as little shape can be derived from only a few frames. As camera frames are added, shape and motion begin to emerge. Though stitching is not performed until the final stage, it is included here in to convey the incremental reconstruction process.

every possible camera and track has been inserted into the reconstruction. Once the spherical model is fully reconstructed, the spherical constraint is removed and bundle adjustment is rerun one final time, allowing full three-dimensional restructuring of the track points (**Figure 10a**). Because each camera and track is well anchored into an overlapping patchwork, unconstrained 3D bundle adjustment progresses toward a more accurate model of the bladder surface.

3.9. Surface Fitting

The reconstruction from section 3.8 yields a set of track points that describe the shape of the bladder. Ultimately, a surface S must be extracted upon which the individual frames can be mapped and stitched together. Though the surface is not constrained to be spherical, S is parameterized in spherical coordinates, such that radius r is given by a function $r = r(\theta, \phi)$. It should be noted that this model implicitly assumes that every point of the bladder surface is visible from the sphere center (**Figure 9a**), and will not suffice in situations where S contains hidden surfaces (**Figure 9b**). Before surface fitting, an initial filtering of the track points is performed to reduce noise from outliers. This is done by median

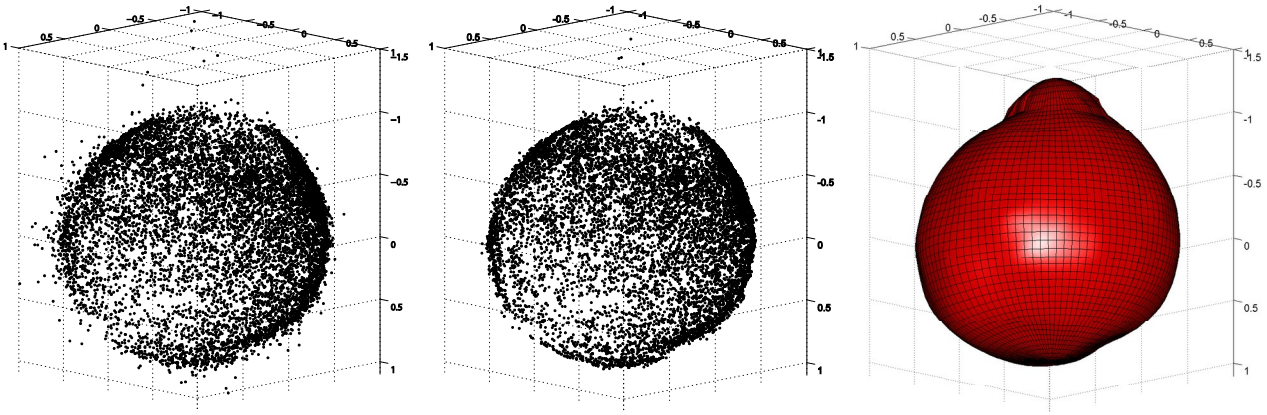
filtering of the radius r of each track's five nearest neighbors in 2D spherical coordinates θ, ϕ (**Figure 10b**).



(a) (b)

Figure 9: The bladder surface S is modeled by a 2D spherical function $r(\theta, \phi)$, in which each surface point is directly visible to the origin at O (a). However, a 2D surface model is insufficient in a case where portions of the surface are hidden (b).

The surface $S$is explicitly defined in a three-dimensional space $S(\theta, \phi, r)$. To construct a surface from the point cloud of tracks points, a smoothing thin-plate spline (TPS) is used to estimate the radius $r(\theta, \phi)$ over spherical grid coordinates (θ, ϕ). Because computation of a TPS increases with the number of points, each grid point of the surface is locally computed using only a subset of track points that spatially inhabit that region of the surface. TPS surface estimation was performed using the MATLAB curve fitting toolbox with a smoothing tolerance of 0.5.



(a) (b) (c)

Figure 10: Unconstrained surface reconstruction of the pig bladder yields a 3D point cloud (a). Noise in the reconstructed track points is mitigated by median radial filtering of nearest-neighbors (b). From the filtered points, a smooth surface contour S is explicitly defined using localized smoothing TPS (c).

3.10. Stitching

The process of stitching involves mapping each video frame onto the composite surface S and subsequently blending images together to create a smooth image of the bladder. First, a 2D panoramic image I_p is initialized, for which each pixel (x_p, y_p) maps linearly to a point in spherical coordinates ¿).

The pixel value $I_p(x_p, y_p)$ is then computed as a weighted contribution of each frame that contains (x_p, y_p), such that:

$$I_p(x_p, y_p) = \frac{\sum_k w(x, y) \cdot I_k(x, y)}{\sum_k w(x, y)}$$

where the mapping between any point (x_p, y_p) in the panorama and a point (x, y) in any image is computed from (2)-(4). Here, we apply center weighting of the image such that $w(x, y)$ is inversely proportional to the distance of (x, y) from the image center (Szeliski and Shum, 1997), thus preferentially weighting the more central regions of each image frame, where there is higher contrast, and surface features are under more direct lighting. For pixels that fall outside the active field of view of the scope, w is set to 0.

4. Results

4.1. Excised Pig Bladder

Over 8,000 video frames comprised the cystoscopic video. From these the software selected 1,061 for analysis. From these a total of 7,797 frame matches were detected between frames, resulting in a total of 10,464 tracks. Through incremental bundle adjustment, a total of 922 frames and 9,954 tracks were inserted into the reconstruction. Frames which did not contain at least 5 tracks and tracks that were not visible to at least 2 cameras were excluded from the model. Following the spherically constrained reconstruction, the mean projection error was 1.92 ± 1.22 pixels. After unconstrained reconstruction the projection error was reduced to 1.66 ± 1.08 pixels. After fitting a smooth surface to the track points, the mean error between the points and surface was 0.02 ±.04. This error is dimensionless for the reason that SfM can only be determined up to a scale factor. Because the bladder is initially constructed on a unit sphere, the radius following unconstrained bundle adjustment is close to one. Thus the error of the surface fit can be interpreted as approximately 2% of the bladder radius (1mm). **Figure 11** presents unfolded views of the 3D stitched surface model of the bladder for each pair of hemispheres. For expert evaluation, the reconstructed bladder can be displayed as single 3D surface in **Figure 13**. The stitched surface further provides validation of a complete scan of the bladder, covering 99.6% of the bladder surface area. Only a small hole is apparent at the neck of the bladder through which the SFE is inserted.

To evaluate the diagnostic quality of the panorama, individual video frames were synthesized by resampling the stitched surface model from the computed camera poses**. Figure 12** compares three example frames take from the raw cystoscopic video compared to the corresponding synthesized views. The synthesized frames appear to be slightly blurred relative to the raw cystoscopic video, diminishing the finer vessel detail. However, lighting artifacts arising from specular reflections of the surface or floating particles are greatly suppressed.

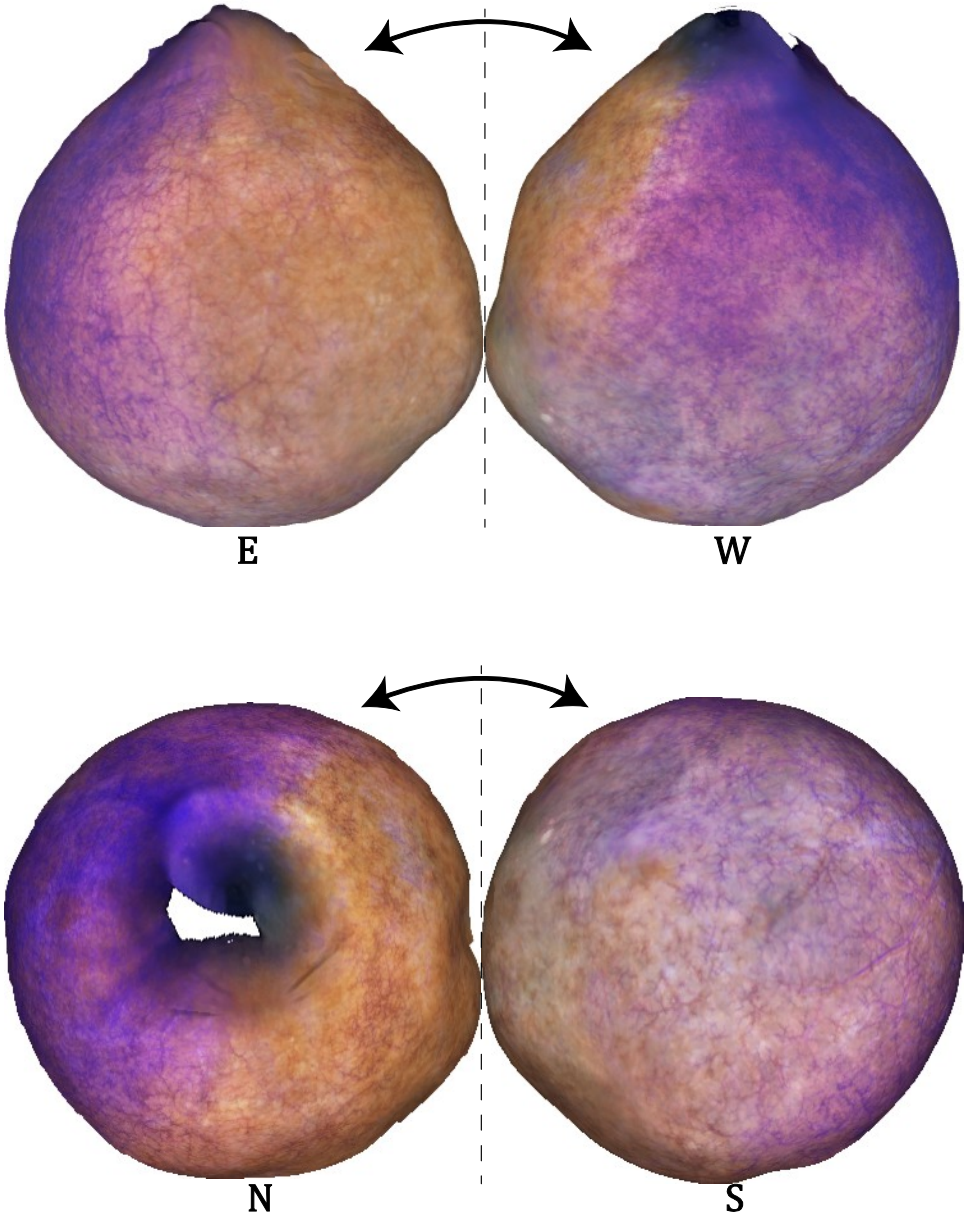


Figure 11: Unfolded views of the 3D stitched surface model of the excised pig bladder by hemisphere.

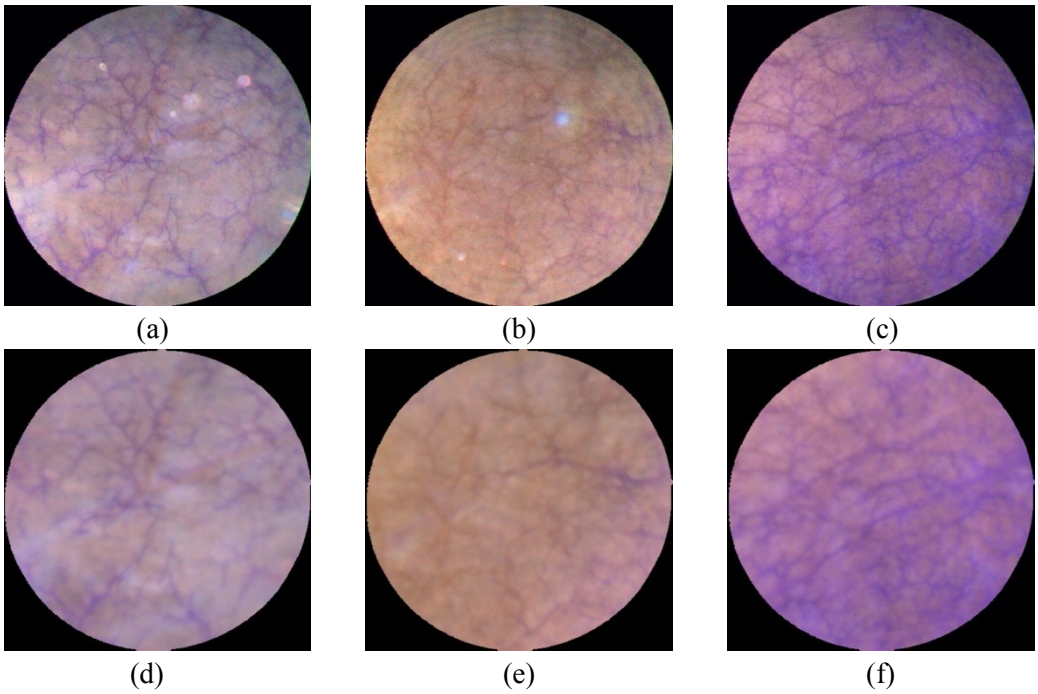


Figure 12: Cystoscopic images (a-c) acquired by the SFE in the excised pig bladder are synthesized from the stitched surface model and prescribed camera locations (d-f). The synthesized frames are blurred, but suppress artifacts related to lighting variation such as specular reflection.

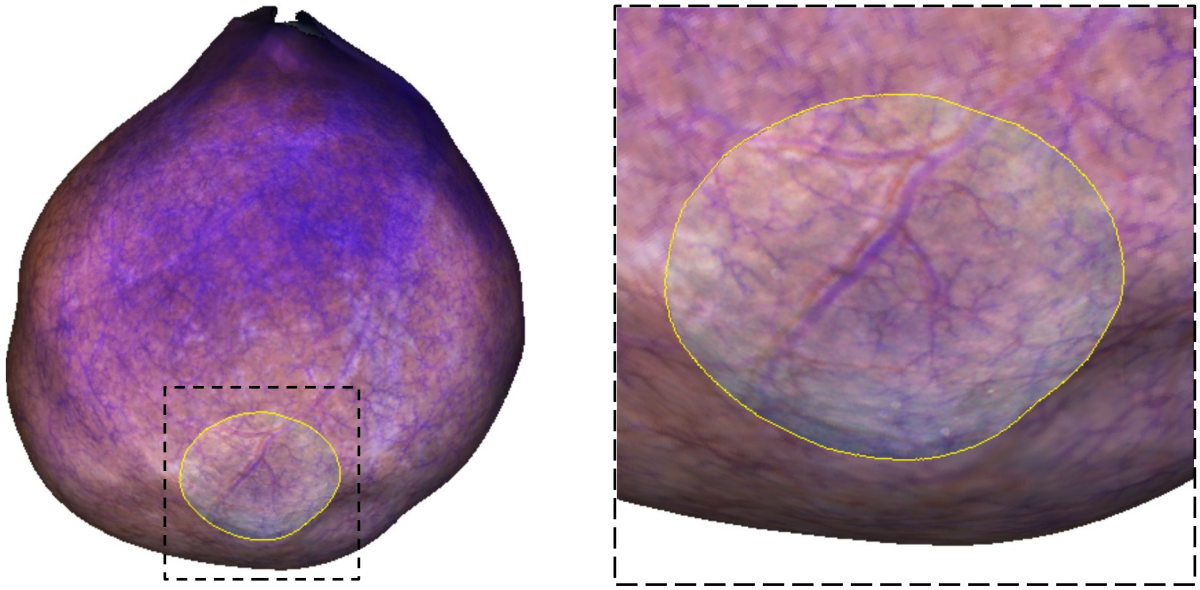


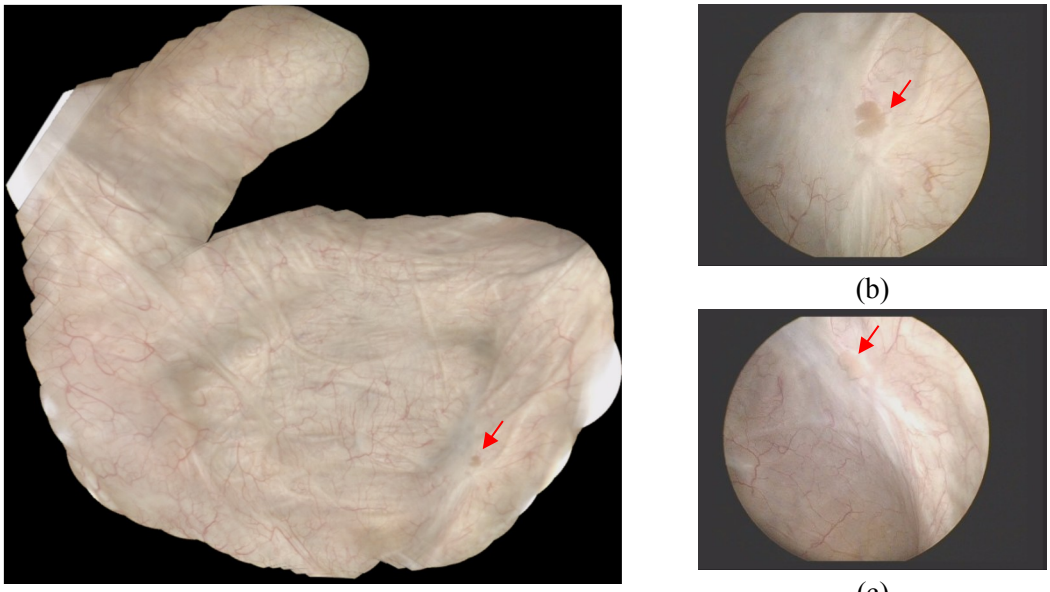
Figure 13: The fully reconstructed 3D model of the excised pig bladder. Individual full resolution video frames can be visualized as overlays on the model and magnified to retain detail for later interpretation by an expert.

4.2. Rigid Cystoscopy

The custom 3D stitching software was tested on cystoscopic video acquired from a routine rigid cystoscopy of a patient at the VA Hospital in Seattle, Washington. The procedure was performed using a

21F rigid cystoscope (Olympus, Tokyo, Japan). High definition video was recorded to the processor unit and converted to an MPEG file. From a video sequence containing 8,275 frames (4.6 minutes), a total of 449 frames were selected. From this, 220 frames were incorporated into the reconstruction (**Figure 14a**). The mean projection error was 4.45 ± 3.09 pixels.

The stitching was fairly accurate and produced a visually well aligned panorama. The somewhat increased pixel projection error over that achieved with the pig bladder experiment is largely owed to the increased image size. Unfortunately, because the rigid cystoscope is rotationally constrained, a full 3D reconstruction was not possible. While the "head on" images of the posterior wall of the bladder could be accurately aligned into the planar reconstruction of **Figure 14a**, matching of frames acquired at glancing angles of the sidewalls was limited. **Figure 14b** depicts a frame that approaches the sidewall of the bladder and which also contains a notable feature correspondence in the panorama. However, as frames are acquired from more glancing angles, such as **Figure 14c**, the frame to frame matching process begins to fail. Two main factors are associated with this breakdown. First, the frame to frame matching process presumes that local alignment of two frames can be achieved by planar homography. However, frames such as **Figure 14c** contain a number of non-coplanar features, and thus do not abide by the local alignment model. Second, frames acquired close to the bladder wall present greater artifact from overexposure, which effects quantitative description of key image features.



(a) (b) (c)

Figure 14: Stitching results from rigid cystoscopic video on a planar map (a). A good reconstruction of the top hemisphere of the bladder is achieved. However, around the fringes, glancing angles of the sidewalls make alignment more difficult and introduce greater artifact caused by overexposure. The red arrow illustrates a shared feature in the panorama and two frames (b&c). While the frame in (b) is incorporated into (a), feature matching fails with the frame in (c) due to the skewed perspective and lighting variation. As a result, it is omitted from the reconstruction.

5. Conclusions

In this work, a novel 3D stitching algorithm is presented that utilizes SfM by bundle adjustment to reconstruct cystoscopic video into a 3D stitched surface model of the bladder. This model is solely acquired from video of the procedure and does not require supplemental image data or foreknowledge of the particular endoscope used or its intrinsic properties. Using cystoscopic video acquired from an excised pig bladder with the SFE, a pixel projection error of less than 2 pixels was achieved, covering 99.6% of the bladder. This demonstrates adequate alignment of images into a single consistent patchwork and the reconstructed shape provides validation that no region was missed. Although there is some loss of feature detail as illustrated in **Figure 12d-f**, visualization could be supplemented with overlays of the native video frames (**Figure 13b&c**). A user interface could allow an urologist to easily switch between dual levels of detail as part of the examination.

While the pixel projection error provides a valuable metric by which to measure consistency of the reconstruction, the lack of a gold standard limits any true evaluation of the reconstruction accuracy. This is due to the fact that the actual bladder shape and endoscope motion are unknown. Ultimately, accurate reconstruction is not assumed to be essential for expert evaluation. Rather, the model serves as visual validation of complete surveillance and permits wide-field examination of the entire bladder. Thus, full 3D reconstruction of the bladder surface is more useful as a means of improving feature alignment and overall stitching quality than divulging actual bladder shape. Despite this, simulated endoscopy of virtual bladder models produced highly accurate reconstructions in previous work (Soper et al., 2011), and may prove to be of value in detecting abnormalities in the surface topology of the bladder. However, unlike the simulated endoscopy experiments, inconsistencies due to lighting, deformation of the bladder, and image distortion may preclude highly accurate surface reconstruction, and should be investigated further.

The software was also tested on video acquired from a patient undergoing rigid cystoscopy. This produced only a partial stitched surface model with a pixel projection error of 4.45. The limited reconstruction is the result of difficulty in matching images of the sidewalls taken from a skewed perspective using our software. Some of these limitations could be overcome by implementing a local 3D alignment algorithm through computation of the fundamental matrix, such as the normalized 8-point algorithm (Hartley, 1997). The only drawback of such a method is that it requires a greater number of correct feature correspondences for RANSAC, which has considerable impact on the overall processing time. However, image artifacts, such as those created by overexposure when the light source is close to the sidewall cannot be eliminated. Furthermore, the motion of a rigid endoscope is largely restricted such that many features are only peripherally visible. Because rigid cystoscopes are not able to retroflex to obtain critical internal views toward the neck of the bladder, image gaps prevent the software from assembling the video into a 3D model or quantitatively ascertaining how much of the bladder is unseen.

Despite this, the matching of frames and subsequent reconstruction is greatly aided by using the SFE as compared to a rigid cystoscope. First, because SFE images are constructed by scanning of focused laser light, overexposure is eliminated so long as some small distance is maintained between the SFE and bladder wall. Second, the scan trajectory achieved using the SFE (**Figure 4c**) permits imaging over a full bend angle of 180°. This allows the SFE to capture images of the surface over a full 360° field of view and at orientations that are nearly normal to the bladder surface.

Panoramic stitching of cystoscopic video is heavily reliant on adequate operation of the endoscope. First, comprehensive coverage of the bladder is essential. In the event that major portions of the bladder are not visualized, many of the reconstructed feature points and camera locations will be under-constrained, leaving large gaps in the model. Second, because reconstruction is driven by matched

features, considerable overlap between frames is critical. Using the spiral scan trajectory in **Figure 4c**, subsequent spirals should ideally overlap by 50%. With diminishing overlap, feature matching becomes increasingly more difficult. This underscores the need for controlled steering of the SFE, which is the subject of an ongoing research project. Instituting machine controlled steering of the SFE would also provide a valuable estimate of the scope pose at each frame, which could greatly diminish time required by the software (currently around 20 minutes) to merely a few minutes.

This initial investigation substantiates use of 3D surface stitching software in conjunction with SFE technology as a potential means of automating bladder cancer surveillance. Further work is required to show that this software, married to our miniature endoscope and automated steering mechanism is sufficiently robust to produce full surface reconstructions under various clinical conditions. Specific challenges include poor visualization due to bleeding or debris in the urine, abnormal bladder shape, or lack of feature detail to instruct matching of frames. Furthermore, a user-interface must be developed to allow the urologist to efficiently navigate the stitched surface model. To guard against failed reconstruction using this software, the native video could be accessible within the user-interface for immediate reference.

6. Acknowledgments

The authors would like to thank Wayne Lamm for his assistance with all *ex vivo* animal work and engineers Rich Johnston, Dave Melville, and Cameron Lee who developed the SFE. This work was supported in part by the W.H. Coulter Foundation, University of Washington Center for Commercialization, and Qatar National Research Fund - NPRP 09-214-2-090 (contract from Qatar University, PI W.J. Yoon).

7. References

Behrens, A., 2008. Creating panoramic images for bladder fluorescence endoscopy. Acta Polytechnica, Czech Technical University in Prague 48, 50-54.

Behrens, A., Heisterklaus, I., Muller, Y., Stehle, T., Gross, S., Aach, T., 2011. 2D and 3D visualization methods of endoscopic panoramic bladder images, In: Kenneth, H.W., David, R.H., III (Eds.). SPIE, p. 796408.

Behrens, A., Stehle, T., Gross, S., Aach, T., 2009. Local and global panoramic imaging for fluorescence bladder endoscopy, 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009, September 2, 2009 - September 6, 2009. IEEE Computer Society, Minneapolis, MN, United states, pp. 6990-6993.

Chang, R.F., Chang-Chien, K.C., Takada, E., Huang, C.S., Chou, Y.H., Kuo, C.M., Chen, J.H., 2010. Rapid image stitching and computer-aided detection for multipass automated breast ultrasound. Med Phys 37, 2063-2073.

Dey, D., Slomka, P.J., Gobbi, D.G., Peters, T.M., 2000. Mixed reality merging of endoscopic images and 3-D surfaces, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2000. Third International Conference. Proceedings, 11-14 Oct. 2000. Springer-Verlag, Berlin, Germany, pp. 796-803.

Engels, C., Stewénius, H., Nistér, D., 2006. Bundle Adjustment Rules, Photogrammetric Computer Vision (PCV '06), Bonn, Germany.

Hartley, R.I., 1997. In defense of the eight-point algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 580-593.

Hernandez-Mier, Y., Blondel, W.C., Daul, C., Wolf, D., Guillemin, F., 2010. Fast construction of panoramic images for cystoscopic exploration. Comput Med Imaging Graph 34, 579-592.

Hu, M., Penney, G., Rueckert, D., Edwards, P., Bello, F., Figl, M., Casula, R., Cen, Y., Liu, J., Miao, Z., Hawkes, D., 2010a. A robust mosaicing method with super-resolution for optical medical images, 5th International Workshop on Medical Imaging and Augmented Reality, MIAR 2010, September 19, 2010 - September 20, 2010. Springer Verlag, Beijing, China, pp. 373-382.

Hu, M.X., Penney, G., Figl, M., Edwards, P., Bello, F., Casula, R., Rueckert, D., Hawkes, D., 2010b. Reconstruction of a 3D surface from video that is robust to missing data and outliers: Application to minimally invasive surgery using stereo and mono endoscopes. in press.

Jemal, A., Siegel, R., Xu, J., Ward, E., 2010. Cancer statistics, 2010. CA Cancer J Clin 60, 277-300.

Lee, C.M., Engelbrecht, C.J., Soper, T.D., Helmchen, F., Seibel, E.J., 2010. Scanning fiber endoscopy with highly flexible, 1 mm catheterscopes for wide-field, full-color imaging. J Biophotonics 3, 385-407.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91-110.

Miranda-Luna, R., Daul, C., Blondel, W.C.P.M., Hernandez-Mier, Y., Wolf, D., Guillemin, F., 2008. Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm. IEEE Transactions on Biomedical Engineering 55, 541-553.

Naya, Y., Nakamura, K., Araki, K., Kawamura, K., Kamijima, S., Imamoto, T., Nihei, N., Suzuki, H., Ichikawa, T., Igarashi, T., 2009. Usefulness of panoramic views for novice surgeons doing retroperitoneal laparoscopic nephrectomy. Int J Urol 16, 177-180.

Olijnyk, S., Hernandez Mier, Y., Blondel, W.C.P.M., Daul, C., Wolf, D., Bourg-Heckly, G., 2007. Combination of panoramic and fluorescence endoscopic images to obtain tumor spatial distribution information useful for bladder cancer detection, Novel Optical Instrumentation for Biomedical Applications III, 17-19 June 2007. SPIE - The International Society for Optical Engineering, USA, pp. 66310-66311.

Quan, L., Lan, Z., 1999. Linear N-point camera pose determination. IEEE Transactions on Pattern Analysis and Machine Intelligence 21, 774-780.

Rai, L., Higgins, W.E., 2009. Fusion of MDCT-Based Endoluminal Renderings and Endoscopic Video, Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging, February 8, 2009 - February 10, 2009. SPIE, Lake Buena Vista, FL, United states.

Seibel, E.J., Carroll, R.E., Dominitz, J.A., Johnston, R.S., Melville, C.D., Lee, C.M., Seitz, S.M., Kimmey, M.B., 2008. Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and Barrett's esophagus. IEEE Transactions on Biomedical Engineering 55, 1032-1042.

Seibel, E.J., Johnston, R.S., Melville, C.D., 2006. A full-color scanning fiber endoscope, Optical Fibers and Sensors for Medical Diagnostics and Treatment Applications VI, 21 Jan. 2006, 1 ed. SPIE - The International Society for Optical Engineering, USA, pp. 608303-608301.

Seshamani, S., Lau, W., Hager, G., 2006. Real-time endoscopic mosaicking, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006. 9th International Conference. Proceedings, Part I, 1-6 Oct. 2006. Springer-Verlag, Berlin, Germany, pp. 355-363.

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. ACM Transactions on Graphics 25, 835-846.

Soper, T.D., Chandler, J.E., Porter, M.P., Seibel, E.J., 2011. Constructing spherical panoramas of a bladder phantom from endoscopic video using bundle adjustment, In: Wong, K.H., Holmes Iii, D.R. (Eds.), 1 ed. SPIE, Lake Buena Vista, Florida, USA, pp. 796417-796412.

Szeliski, R., Shum, H.-Y., 1997. Creating full view panoramic image mosaics and environment maps, Proceedings of the 1997 Conference on Computer Graphics, SIGGRAPH, August 3, 1997 - August 8, 1997. ACM, Los Angeles, CA, USA, pp. 251-258.

Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 1999. Bundle adjustment-a modern synthesis. Springer, Berlin, Germany, pp. 298-372.

Yaniv, Z., Joskowicz, L., 2004. Long bone panoramas from fluoroscopic X-ray images. IEEE Trans Med Imaging 23, 26-35.

Yoon, W.J., Park, S., Reinhall, P.G., Seibel, E.J., 2009. Development of an Automated Steering Mechanism for Bladder Urothelium Surveillance. J Med Device 3, 11004.