8

The STIFF-FLOP Vision System

Erwin Gerz, Matthias Mende and Hubert Roth

Lehrstuhl für Regelungs- und Steuerungstechnik (RST), Department Elektrotechnik und Informatik, Fakultät IV – Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Siegen, Germany

Abstract

The use of new soft robots in minimally invasive surgery offers exciting new possibilities while it generates new challenges for the technical implementation. This chapter presents methods for the detection of the STIFF-FLOP arm using visual sensing means. Based on the image information of an endoscopic camera, the visible sections are evaluated to determine the position of the manipulator. A variety of algorithms for the detection of the STIFF-FLOP arm as well as for the detection of its module connectors will be described.

A stereo camera is used to register all components in a common frame. A transformation tree is set up to refer the position of the STIFF-FLOP arm in the endoscopic camera image to the base of the STIFF-FLOP arm.

All methods have been integrated and tested in the newly developed system.

8.1 Introduction

To monitor and to control the STIFF-FLOP arm, the video data of the surgeon's endoscope is processed and evaluated. The biggest challenge lies in the reliable detection of the STIFF-FLOP arm. The nature of the object itself and the conditions in the workspace as well as the equipment available for minimally invasive surgery (MIS) leads to several restrictions in the implementation strategies.

Established methods are not applicable in the examined scenario, as these are generally either based on the detection of a known and trained outline, a clear, well-known texture or on the pronounced differences in contrast of foreground and background. None of these conditions are given here: The manipulator is flexible throughout and can change its shape and length. In addition, texture recognition is difficult to implement, since any fitted pattern is distorted significantly with increasing curvature of the arm.

The challenge here is in the detection of a flexible arm, capable of changing its shape and size during the movement, as well as the fact that the visible section of the arm can vary in the endoscope image. With the aim to develop a high-performance automated learning and recognition method, a two-step algorithm has been designed. On the one hand a texture-based pattern recognition and classification method based on Support Vector Machines (SVMs) [1–3] has been implemented [4]. The second step is the detection of optical circular markers with a modified circle detection algorithm.

8.2 Optical Tracking of the STIFF-FLOP Arm

The vision system for the tracking of the STIFF-FLOP arm consists of a 3D-tracking system (Axios, Cambar B2, Germany) and two endoscopic camera systems (Richard Wolf, Endocam 5509 and Richard Wolf, Endocam Performance HD, Germany).

In order to process the image data of the endoscopic camera, the video is streamed to ROS using a frame grabber (Intensity Pro, Blackmagic, Australia).

In the first step the endoscopic cameras will monitor and detect the STIFF-FLOP arm. In the second step the 3D-tracking system will track locators mounted on the endoscopic camera systems. Based on this data a transformation tree is set up which allows registering the position of the endoscopic camera to the base of the STIFF-FLOP arm. Figure 8.1 shows the setup for the operating room.

8.2.1 Axios Measurement System Cambar B2

The tracking system Cambar B2 is a stereo imaging system for highly accurate measurement of 3D coordinates of signalized points within a specified measurement volume. It consists of hardware components as well as software parts controlling the system.

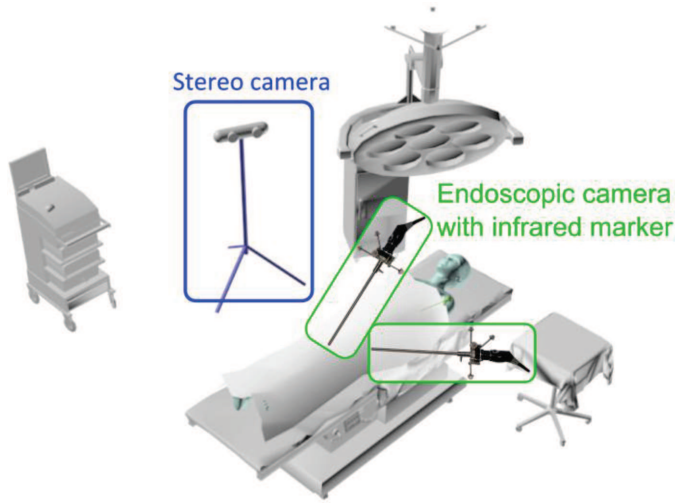

Figure 8.1 The endoscopic camera is used to detect the STIFF-FLOP arm at its destination while it is monitored by the stereo camera.

Figure 8.2 Stereo camera Axios Cambar B2, Axios 3D, Oldenburg, Germany.

The Cambar imaging system, which is shown in Figure 8.2, detects and measures passive, i.e., retroreflective points. These markers are either measured as single points or – if they fulfill a pre-defined marker geometry – as locators or rigid bodies. Measured points are classified according to their image characteristics and geometry. Afterwards, they are assigned to accuracy classes to describe and evaluate their influence on the maximum achievable measurement accuracy. The locators should meet several requirements that are summarized in Table 8.1.

Table 8.1 Specifications for the locator design

	Minimum Requirements	Recommended Specifications
Minimum number of markers	3	$\geq$4
Marker shape and marker surface	Sphere, dot (flat circle) (retro-reflective)	Sphere (retro-reflective)
Marker diameter	10–12 mm (depends on distance from camera to measured object)	10 mm
Minimum distance between marker centers	Twice the minimum marker diameter	Twice the actual marker diameter
Distance of segments between markers to other segments within a locator.	> Minimum diameter	> Marker diameter
Requirements regarding rigid body geometry	Points must not be aligned in a straight line.	Points should be spread-out in space (in *x*-, *y*-, and *z*-direction), not be on the same plane.

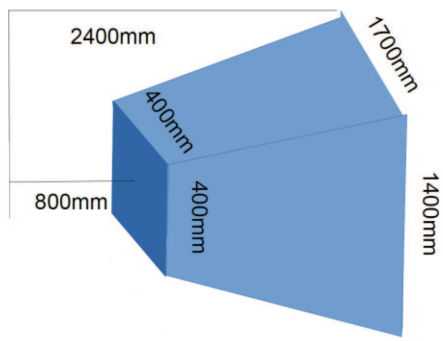

Figure 8.3 Pre-defined measurement volume of the Axios camera Cambar B2 [5].

It is advisable to use markers of the recommended specifications in order to achieve maximum accuracy. The system is capable of tracking and measuring points in a pre-defined measurement volume, which is displayed in Figure 8.3.

8.2.2 The Endoscopic Camera System

In order to find and to prove the pose and orientation of the developed STIFF-FLOP manipulator, two endoscopic cameras were ordered as shown in Figure 8.4. Each camera system consists of a light module with a xenon

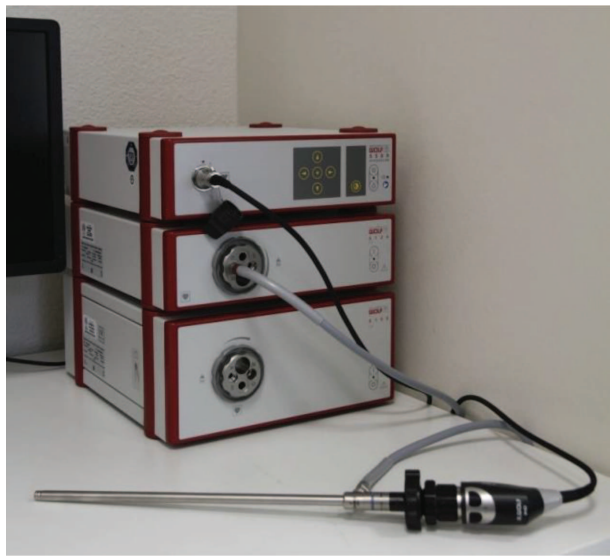

Figure 8.4 The endoscopic camera system.

light source (Light Projector 5124 and 5132, Richard Wolf, Germany) and a camera module (HD Endocam 5509 and Endocam performance HD, Richard Wolf, Germany), which is the interface to the camera on the laparoscope. The laparoscope connects the lens and the camera as well as the fiber of the light source.

The aim is to observe the STIFF-FLOP arm while it is operated by the surgeon. The laparoscopic cameras provide a video stream in full HD resolution (1920*1080, @50 Hz), in medical HD resolution (1280*1024, @50 Hz) or in HD ready resolution (1280*720, @50 Hz). To enable clinical use, all parts of the system are autoclavable (except the disposable reflectors). The video streams will be analyzed using image processing algorithms in order to detect the STIFF-FLOP arm.

8.2.3 Image Processing on Endoscopic Camera Images

The video stream of the endoscopic camera is captured with a frame grabber and streamed to ROS afterwards. Here the integrated image processing functions can be used to calibrate the camera and correct the image distortion which is displayed in Figure 8.5.

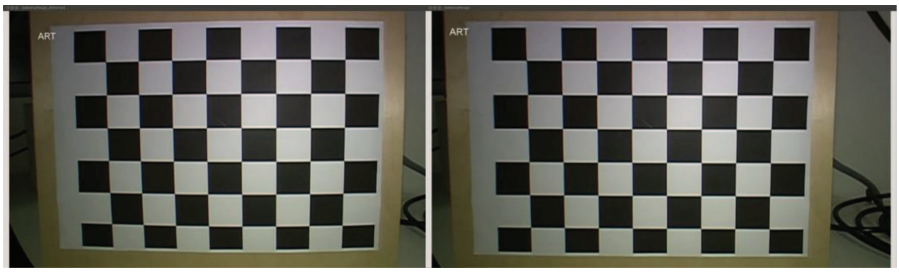

Figure 8.5 The left image shows the original image, the right shows the undistorted image.

8.2.3.1 Removal of specular reflections

Specular reflections occur if light shines on a surface. According to the laws of reflection, the reflected light beam has the same angle as the incident light beam relative to the normal of the surface. Based on the physical structure of the endoscopic cameras it is nearly impossible to get an image without specularities. These areas of overexposure lead to a corrupted filter mask, so that the algorithm could not detect the observed structure reliably. In order to remove these highlights and to reconstruct the original color, different methods were compared.

Shen and Cai [6] introduced an effective method to separate specular reflections and diffuse reflection components in multi-colored textured surfaces using a single image.

The first procedure in this method includes scanning the image to determine the minimum value of RGB components of each pixel. Each minimum value is subtracted from all three RGB components in the corresponding pixel to produce the specular-free image. After that, a threshold is added to each pixel to compensate the loss of the chromaticity that occurred because of the subtraction performed earlier to produce what Shen and Cai called the *Modified Specular Free Image*.

Another promising method was described by Miyazaki et al. [7], which is applied on single images. It does not apply any region segmentation or consider any relations between the neighbor pixels. This makes the execution time dependent only on the size of the image. The geometry of textures in the image is maintained and it does not affect the execution time either. The hue and saturation of the image do not change after the process, but the intensity does. The color changes slightly as well, but it remains similar to the color in the original image.

The method obtains the specular-free image by transforming the original image from its RGB color space into another customized color space introduced by Miayazki et al. In that color space, a filter is applied on the image data to eliminate the specular reflection components. The image is then transferred back into the RGB color space.

Both methods were implemented and tested. A comparison between the images with highlights and the result after applying both methods are shown in Figures 8.6 to 8.8.

In different tests, the method described by Miyazaki performed about 20% faster than the method described by Shen and Cai. By applying these methods on the video stream, large specular reflections were removed completely in almost all procedures. Restoring the original surface color behind large specular reflections is still problematic and as a result, those areas appear grey after the specular reflection removal was applied.

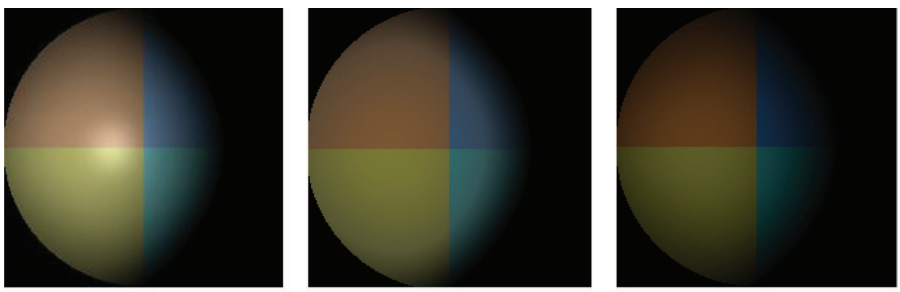

Figure 8.6 (a) Sphere with four different colors and a spot of highlights almost in the middle; (b) the result image after applying Shen and Cai's method; and (c) the result image after applying Miyazaki's method.

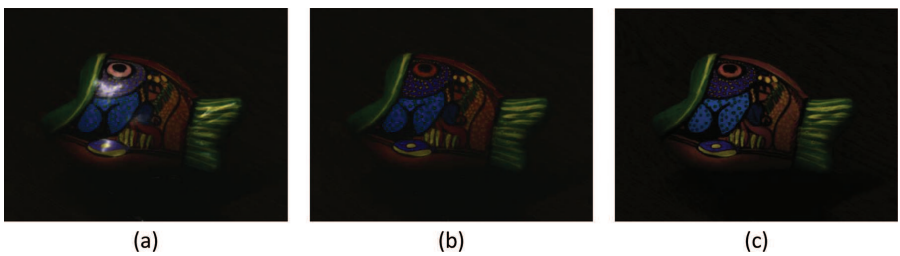

Figure 8.7 (a) Highlights on a fish; (b) the result image after applying Shen and Cai's method, and (c) the result image after applying Miyazaki's method.

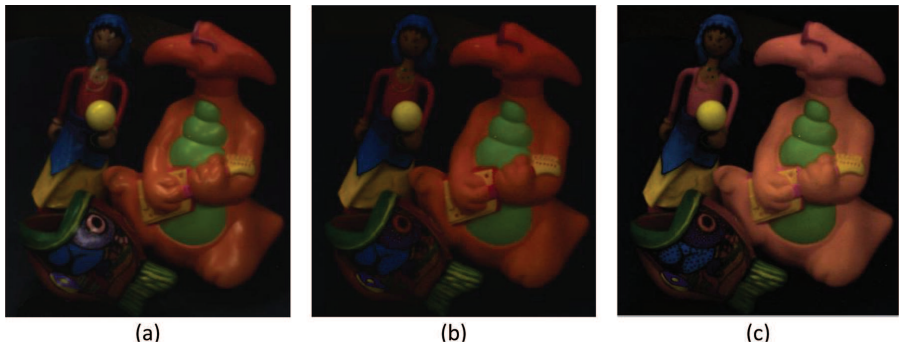

Figure 8.8 (a) Highlights on various toys; (b) the result image after applying Shen and Cai's method, and (c) the result image after applying Miyazaki's method.

8.2.3.2 Improvement of the dynamic range

Another approach that became popular in the last years is to extend the dynamic range of images. Typically this procedure requires a huge number of computations, so it is not applicable for real-time applications.

Inspired by this approach, a light-weight high dynamic range method was implemented. Hereby, one image taken with a long exposure time is combined with an image taken with a short exposure time (Figure 8.9).

After converting both images to an HSV (hue, saturation, value) color space, white areas of the lighter image are detected. These areas are also identified in the darker image, where the brightness value of the darkest pixel is subtracted from all pixels of the image. In a second step, the brightness channel of both images is added up to a new image. This newly generated image unfortunately exceeds the limits of HSV specifications, so the brightness layer has to be shrunk back into valid borders.

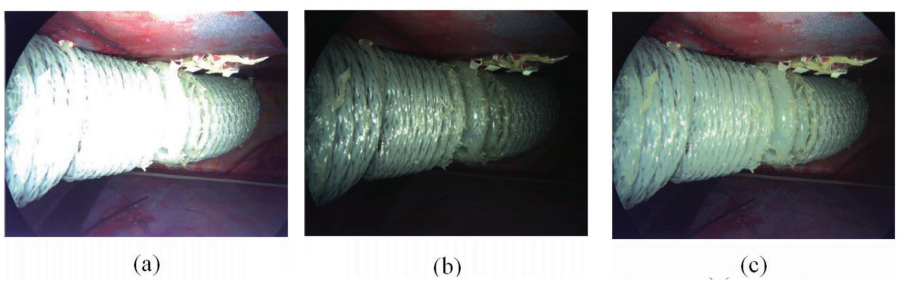

Figure 8.9 (a) Image taken with long exposure time, (b) image taken with short exposure time, and (c) the combined image with an improved dynamic range.

The resulting data works quite well as an input stream for the object-detection algorithms. Unfortunately switching from one aperture to another causes an unexpected delay, so the resulting frame rate is not satisfying.

8.2.4 Detection of the STIFF-FLOP Arm in the Camera Image using Machine Learning Algorithms

In this section, the detection of the STIFF-FLOP arm in the endoscopic camera image is described. In the first approach, it was assumed to color-code the STIFF-FLOP manipulator in an equal color and apply a color filter to detect it. Following this approach it was possible to detect the colored arm. But as soon as the color on the STIFF-FLOP arm would become contaminated the detection would fail. In the second approach, a machine learning algorithm was used to detect the STIFF-FLOP arm.

The methodology which provided the most reliable results is based on the usage of SVMs. For the use of those, a set of training samples is needed, where the searched object (i.e., the STIFF-FLOP arm), as well as the background are labeled correspondingly. The model of SVM represents the samples as points in space. They are mapped in a way that the categories are separated by a gap. For the largest gap, the recognition will provide the most stable results. This idea of data classification can be applied to images. By analyzing images, they can be classified into categories.

For this application the idea is to divide the camera image into small squares, sized about 25 $\times$ 25 pixels, which are analyzed and classified as background or as object (i.e., the STIFF-FLOP manipulator), labeled with the variable $y_i \in \{-1,\ 1\}$.

$$D = \{(x_i, y_i)\,|x_i \in \Re^p,\ y_i \in \{-1,\ 1\}\}_{i=1}^{n} \tag{8.1}$$

where D represents the number n of quadratic parts x_i of the initial image. In Figure 8.10 a simplified visualization of the previous equation is shown, whereby the red squares represent one classification and the blue circles represent the other classification (i.e., background and object).

Assuming a dataset as observed in Figure 8.10, a hyperplane can be defined which separates both classes. The green line in Figure 8.10 represents an optimal hyperplane, which separates both classes with the maximum margin. This hyperplane will act as a classifier to recognize the object in the camera image.

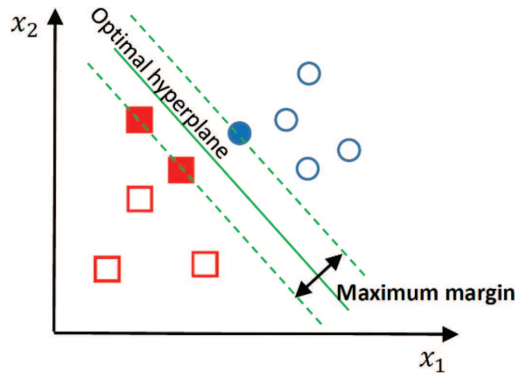

Figure 8.10 Here the different parts of the image are visualized according to their characteristics, whereby the red squares represent the first class and the blue circles the second class.

Following the procedure described by [8] and [4], a hyperplane can be constructed as the set of points $\underline{x}$ satisfying the equation:

$$\underline{w}' \cdot \underline{x} + b = 0 \tag{8.2}$$

Where $\underline{w}$ is the normal vector of the hyperplane and b is the bias. The sample points can be either found above the upper or below the lower side of the hyperplane:

$$\begin{aligned} \underline{w}' \cdot \underline{x} + b \geq 1 \quad for \quad y_i = +1 \\ \underline{w}' \cdot \underline{x} + b \leq 1 \quad for \quad y_i = -1 \end{aligned} \tag{8.3}$$

The samples on the upper and lower margin are the support vectors. Considering y_i they are defined as:

$$y_i(\underline{w}' \cdot \underline{x} + b) - 1 = 0 \tag{8.4}$$

In order to find the maximum margin, the distance between the support vectors on the upper and lower border has to be calculated.

The points on the border are defined as:

$$\begin{aligned} x_+ &= \frac{1-b}{\underline{w}} \quad for \quad y_i = +1 \\ x_- &= \frac{1-b}{\underline{w}} \quad for \quad y_i = -1 \end{aligned} \tag{8.5}$$

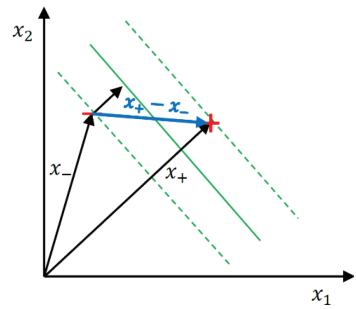

Figure 8.11 The distance between the point x_- on the lower margin and the point x_+ on the upper border is visualized.

Considering the geometry, the width is calculated by subtracting the points on the border, as displayed in Figure 8.11.

$$width = (x_+ - x_-) \cdot \frac{\underline{w}}{\|\underline{w}\|} = \frac{2}{\|\underline{w}\|} \tag{8.6}$$

In order to maximize the margin, the minimum of $\underline{w}$ has to be determined. Therefore, the minimum of $\underline{w}$ can be substituted:

$$substitute: \quad \min \|\underline{w}\| \quad by \quad \min \tfrac{1}{2} \|\underline{w}\|^2 \tag{8.7}$$

In order to minimize $\underline{w}$ the argument is extended with the support vectors $y_i(\underline{w}' \cdot \underline{x} + b) - 1 = 0$ multiplied by the Lagrangian α:

$$L = \arg \min_{w,b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|\underline{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i(\underline{w}' \cdot \underline{x} + b) - 1 \right] \right\} \tag{8.8}$$

The minimum is found by setting the first derivative to zero. The resulting $\underline{w}$ is found as a linear combination of the samples with $\underline{x}_i$ as support vector.

$$\begin{aligned} \underline{w} &= \sum_{i=1}^{n} \alpha_i y_i \underline{x}_i \\ \sum_{i=1}^{n} \alpha_i y_i &= 0 \end{aligned} \tag{8.9}$$

If the classes cannot be separated linearly, the argument can be extended by a polynomial or a radial basis function, allowing a more complex separation [2]. For this application, the best results were obtained with a polynomial approach of third order.

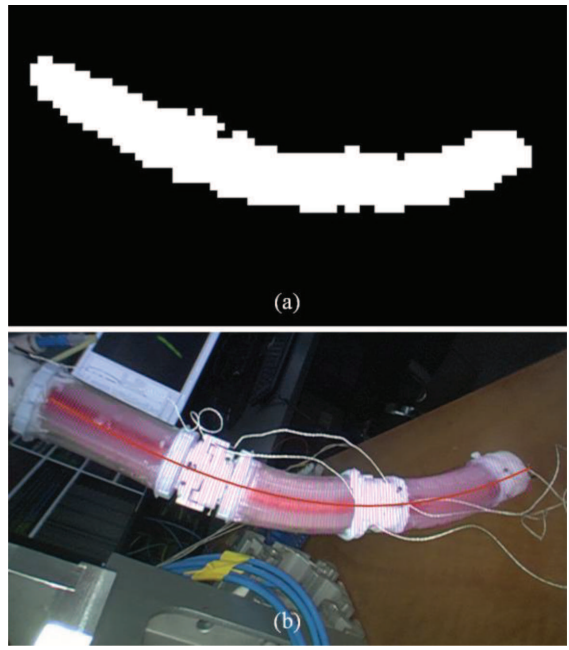

Figure 8.12 (a) Detected parts of the STIFF-FLOP arm using the prior calculated hyperplane (b) corresponding camera image with the highlighted centerline (red) and perpendicular to it the diameter is calculated (violet).

After the hyperplane is identified, it can be used for a real-time application providing the possibility to process the acquired camera images directly. During the real-time application, the same idea is pursued. The camera image is grained into small parts with the same size that was used during the learning process. Afterwards these small parts are classified using the calculated hyperplane. As result, the detected contour is visualized in Figure 8.12.

The next step is to calculate the spatial position of the manipulator in the camera coordinates. Therefore, the center line of the detected STIFF-FLOP arm is determined and the diameter is calculated accordingly. Based on the information, the planar coordinates (i.e., x and y) can be extracted from the pixel coordinates and the distance of the STIFF-FLOP arm to the camera's optical center is calculated based on the determined diameter.

In Figure 8.12, the usage of the implemented machine learning STIFF-FLOP arm detection algorithm is demonstrated, allowing an estimation of the position of the STIFF-FLOP arm in camera coordinates.

8.2.5 Detection of the Module Connection Points of the STIFF-FLOP Arm

To achieve a reliable and robust tracking result, in particular when the STIFF-FLOP arm is being maneuvered in front of unknown and challenging backgrounds inside the abdominal area, unambiguous green ring markers are placed on the STIFF-FLOP manipulator at distinct locations as shown in Figure 8.13.

The color is chosen due to the high contrast with respect to the expected color scheme (mainly in the red spectrum) during surgical interventions inside the abdomen. The ring markers are fixed on the force/torque sensor structures that are placed between segments and at the end of the STIFF-FLOP manipulator.

The detection of these circular markers has several advantages. Firstly, the figure remains almost distortion-free with sufficiently narrow markings. On the other hand, a possibly occurring system-related radial extension of the arm solely affects the accuracy of the distance measurement between the camera plane and the central axis of the manipulator. The presented method is also robust against noise or smaller highlights, as long as the contour of the ring is not interrupted over larger sections.

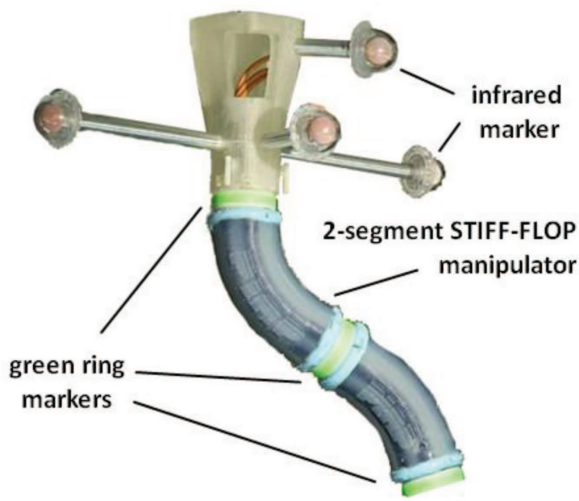

Figure 8.13 Two-segment STIFF-FLOP manipulator with green marker rings placed at the base of the STIFF-FLOP arm and at the module connectors.

Reproducible results were achieved by the application of green ring markers as displayed in Figure 8.13. The idea is based on an algorithm for camera calibration with two arbitrary coplanar circles [9].

The detection process is visualized in Figure 8.14. As a first step, a mask containing all the green colored areas of the video image is generated (2). This mask is the source for a standard contour detection algorithm. The detected contours are filtered based on a set of rules eliminating noise and other unsuitable objects so that only sections of ring-like objects of a suitable size should reach the next step (3).

The remaining contours are split into a concave (4, blue curve) and a convex (4, red curve) part. Each of them is now treated like the visible part

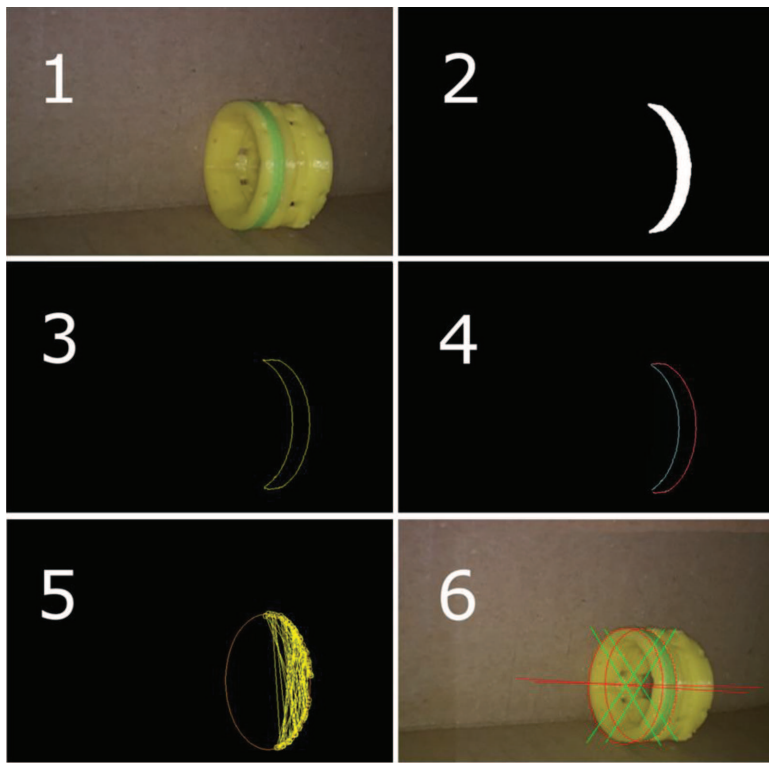

Figure 8.14 Ring marker detection sequence, (1) shows the camera image of the ring marker on the module connector, (2) visualizes the detected mask of the ring marker, (3) highlights the contour of the mask, in (4) this contour is divided into a concave (blue) and in a convex (red) part of an ellipse. In (5) the RANSAC approach to fit an ellipse into the detected shape is demonstrated, while (6) shows the finally determined ellipses with their center positions.

of a closed elliptical shape. Unfortunately, these contours cannot be used directly to determine the parameters of an ellipse, because they contain too much noise and therefore the ellipse parameters are oscillating in a wide range.

A RANSAC [10] approach seems suitable to improve the results, where 10% of each contour's points are selected and fed into a least-square-error algorithm to define the ellipse's parameters. The resulting ellipse is compared with the contour as a whole and the average error is evaluated. If the ellipse fits well, the parameters are stored. If not, another set of points is generated as a basis for the parameter calculation. The results can be seen in (5).

The determined ellipses can be seen as projections of circles. Knowing the circles' diameters, the calculation of those circles' poses in space is possible (6). As an ellipse is effectively a projection of a circle, the position of the midpoint can be determined which is equivalent with the center position of the force/torque sensor. The green ring markers are detected in the camera's coordinate system and are transformed in the manipulator's base system.

8.2.6 Registration of the Endoscopic Camera Image to the STIFF-FLOP Arm

In order to connect the image plane of the endoscopic camera with the world coordinates of the STIFF-FLOP arm, both the endoscopic camera and the base of the STIFF-FLOP arm were equipped with infrared markers, which are tracked by an optical 3D localization system (Axios 3D, Cambar B2, Germany). Using this, a transformation tree was determined to register the image plane of the endoscopic camera in the world coordinates of the STIFF-FLOP arm.

The ring markers are detected in the camera's coordinate system and have to be transformed in the manipulator's base system. Therefore the camera coordinate system has to be integrated into the transformation tree of the whole system by using the output of the Cambar B2 tracking system. An overview of the complete transformation tree can be seen in Figure 8.15.

The transformation tree shows that the world coordinates of the STIFF-FLOP arm are connected to the STIFF-FLOP base (SF_BASE). Through the stereo camera, the transformation of the endoscopic camera (ENDO) is known and connected to the STIFF-FLOP base.

The position of the detected ring markers are known in the frame of the endoscopic camera and are connected below in the transformation

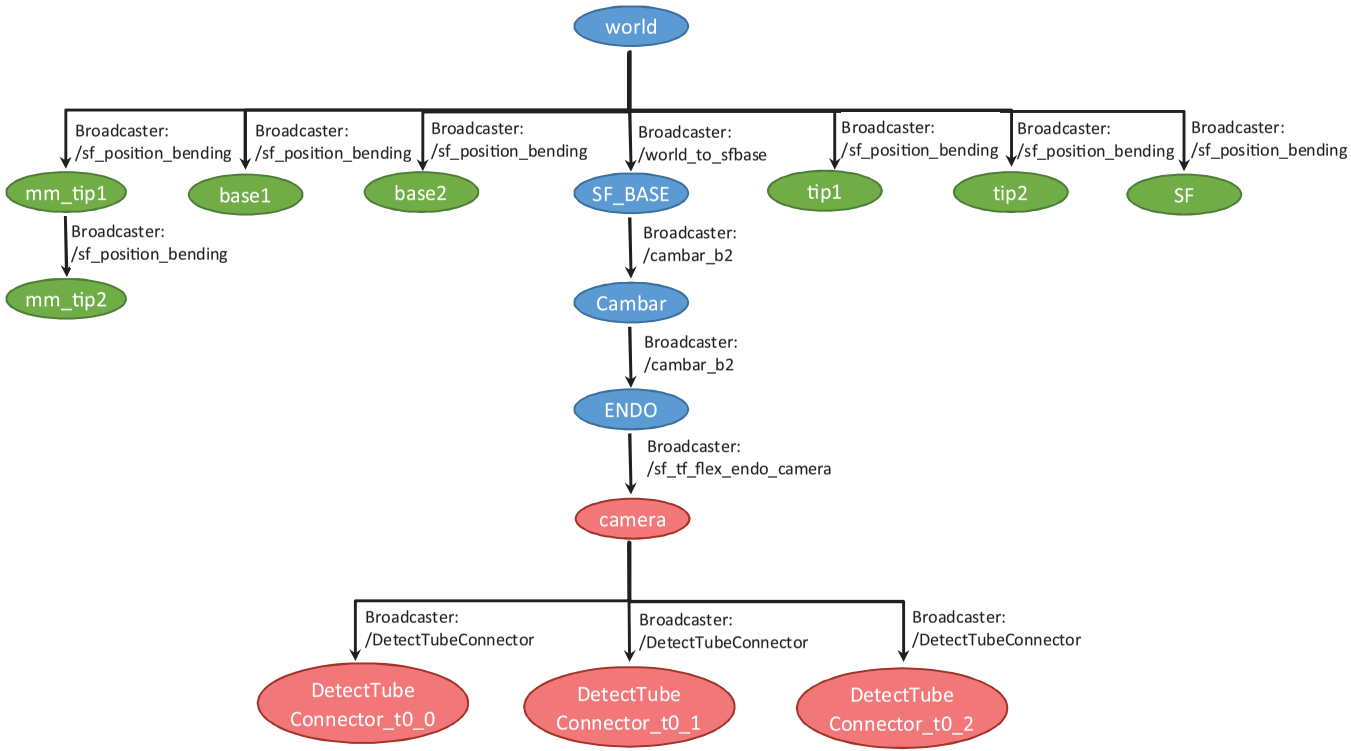

Figure 8.15 Illustration of the transformation tree of the STIFF-FLOP setup connected through ROS. The transformations generated by the optical 3D Tracking system are marked blue, and the transformations resulting from the vision system's output are marked red. The ovals describe the different Frames. The black arrows show through which Broadcasters the Frames are connected.

tree. Thereby the transformation of each detected position relative to the STIFF-FLOP base is known.

8.3 Integration and Validation of the Implemented Methods

Both the calculated centerline and the identified module connection points were determined in the image plane of the endoscopic camera. Using the transformation tree, those positions were transferred to the world coordinate frame of the STIFF-FLOP arm.

The vision system takes over the function of a second-level observation system – therefore an independent ROS node with an alert function was implemented. This node compares the connector's frames sent by the control algorithm with the detected frames published by the vision system and calculates a normalized warning level (0.0–1.0, the absolute value for inacceptable displacements of the manipulator can be reconfigured dynamically). The integration of the vision system into the open loop control is visualized in Figure 8.16.

The position of the used coordinate systems is visualized in the 3D visualization tool for ROS (rviz) (Figure 8.16). The vision system supplies a second-level observation system and optimizes the position data obtained

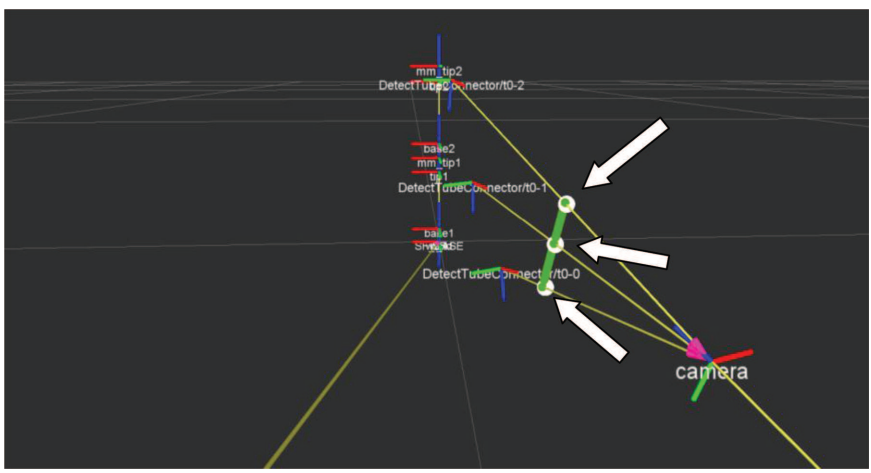

Figure 8.16 Visualization of the transformation tree, showing the matching of the detected positions of the ring markers (white bulbs) with the position provided by the control system (green line connections between ring connections).

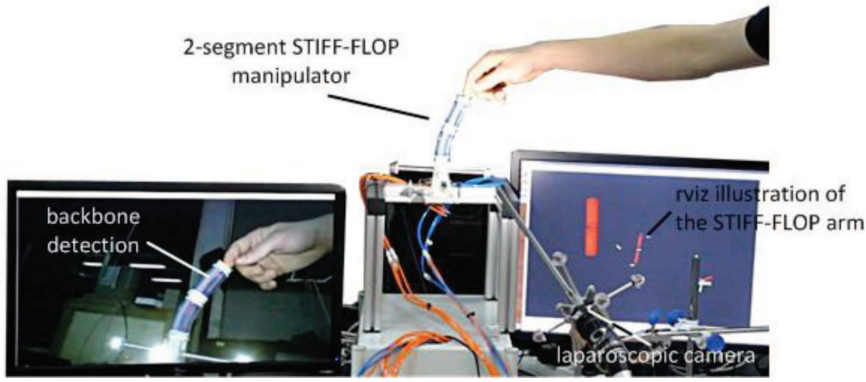

Figure 8.17 Experimental setup, demonstrating the integration of the implemented method, by detecting a deviation between the position detected by the vision system (real arm in the center and the corresponding image on the left screen) and the position of the model of the STIFF-FLOP arm (right screen) which is visualized using the ROS visualization tool (rviz). The deviation is highlighted by coloring the arm in red (right screen).

from calculations using force/torque sensor data and bending sensor information. Figure 8.17 shows the experimental setup: A 2-segment STIFF-FLOP manipulator is mounted on a rack with three green ring markers attached to the base of the robot arm and at the tip of each segment. On the left screen, the video screen of the endoscopic camera is shown. The previously described image processing algorithm detects the markers. A spline interpolation (red) exposes the backbone of the manipulator. The right screen shows this data within the 3D visualization tool.

Both algorithms have been verified in an experimental setup, which is shown in Figure 8.17. The manipulator is located in the middle; the left monitor shows the recorded video stream. The right monitor shows a simulation of the manipulator based upon the model that is used for controlling the manipulator.

The detected position of the connectors is compared to the positions of the control model. If a deviation – which might be caused by external forces – is detected, the deformation is recognized by the optical tracking system and an alarm signal is displayed by coloring the simulated arm in red.

A comparison of the detected positions of the STIFF-FLOP arm and the location given by the control model is possible after transforming all data into a common frame. Afterwards an evaluation of the detection error is possible and the control parameters can be adjusted accordingly.

8.4 Conclusion

The detection of these circular markers has several advantages. Firstly, the figure remains almost distortion-free with sufficiently narrow markings; on the other hand, a possibly occurring system-related radial extension of the arm solely affects the accuracy of the distance measurement between the camera plane and the central axis of the manipulator. The presented method is also robust against noise or smaller highlights, as long as the contour of the ring is not interrupted over larger sections.

The detection of the STIFF-FLOP arm with the implemented SVM method allows a trainable approach to adapt the detection of the STIFF-FLOP arm to the given scenario.

The redundancy of the two methods used (SVM and Ring Marker Detection) essentially offers two advantages: First, it allows a plausibility check of the detected marker positions; on the other hand the calculation of a continuous center line is possible. In addition, this center line can be used for a simple collision detection, if 3D models of the working space are provided.

Acknowledgements

The work described in this paper is funded by the Seventh Framework Programme of the European Commission under grant agreement 287728 in the framework of EU project STIFF-FLOP.

References

[1] Vapnik, W., and Chervonenkis, A. (1974). *Theory of Pattern Recognition* [in Russian] (USSR: Nauka).

[2] Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, PA, 144–152.

[3] Winston, P. H. *6.034 Artificial Intelligence: Support Vector Machines, Massachusetts Institute of Technology: MIT OpenCourseWare*. Available at: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010

[4] Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2:27.

[5] Axios 3D services (2014). *Optisches Messsystem CamBar B2 C8, Datasheet*, *Mai 2014*.

[6] Shen, H. L., and Cai, Q. Y. (2009). Simple and efficient method for specularity removal in an image. *Appl. Opt*. 48, 2711–2719.

[7] Miyazaki, D. (2013). Specular-Free Image. Hiroshima: Hiroshima City University. Available at: http://www.cg.info.hiroshima-cu.ac.jp/∼miyazaki/knowledge/teche40.html

[8] Ben-Hur, A., and Weston, J. (2010). *A User's Guide to Support Vector Machines, Methods in Molecular Biology.* Clifton, NJ: Humana Press. Available at: https://www.researchgate.net/publication/41896604

[9] Chen, Q., Wu, H., and Wada, T. (2004). *Camera calibration with two arbitrary coplanar circles. Computer Vision—ECCV 2004. Lecture Notes in Computer Science*, Vol. 3023 (Berlin, Heidelberg: Springer), 521–532.

[10] Fischler M. A., and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*. 24, 381–395.