3D Position Tracking System for Flexible Cystoscopy

Munehiro Nakamura *m-nakamura@blitz.ec.t.kanazawa-u.ac.jp*
Department of Natural Science and Engineering
Kanazawa University
Kanazawa, 9201192, Japan

Yusuke Kajiwara *kajiwara@de.is.ritsumei.ac.jp*
Department of Information Science
Ritsumeikan University
Kusatsu, 525877, Japan

Tatsuhito Hasegawa *t-hasegawa@blitz.ec.t.kanazawa-u.ac.jp*
Department of Natural Science and Engineering
Kanazawa University
Kanazawa, 9201192, Japan

Haruhiko Kimura *kimura@blitz.ec.t.kanazawa-u.ac.jp*
Department of Natural Science and Engineering
Kanazawa University
Kanazawa, 9201192, Japan

Abstract

Flexible cystoscopy is an examination that allows physicians to look inside the bladder. In flexible cystoscopy, beginner physicians tend to lose track of the observation due to complex handling patterns of a flexible cystoscope and poor characteristics of the bladder. In this paper, as a diagnostic support tool for beginner physicians in flexible cystoscopy, we propose a system for tracking the observation using cystoscopic images. Our system discriminates three handling patterns of a flexible cystoscope, namely bending, rotation, or insertion. To discriminate the handling patterns accurately, we propose to use the degree of bending, rotation, or insertion as features for the discrimination as well as ZNCC-based optical flows. These features are learned by a Random Forest classifier. The classifier discriminates sequential handling patterns of the cystoscope by a time-series analysis. Experimental results on ten videos obtained in flexible cystoscopy show that each of the three handling patterns were correctly discriminated over 90% in average. In addition, we reproduced the observation in a virtual bladder we propose.

Keywords: Flexible Cystoscopy, Position Tracking, Optical Flow, Zero-mean Normalized Cross-Correlation, Handling Pattern.

1. INTRODUCTION

With increase of aged people in the world, incidents of bladder disease are gradually increasing[1]. Bladder disease can be detected in cystoscopy or non-invasive examinations such as blood test, MRI, CT, PET, and ultrasonography. The non-invasive examinations are painless and less stressful. However, it is still difficult to detect tiny legions in a non-invasive imaging examination[2]. Cystoscopy is conducted when a lesion found in a non-invasive imaging examination or sever symptoms are appeared in a patient. Cystoscopy enables physicians to look inside the bladder to confirm a patient's legion directly. There are two types of cystoscopes, namely rigid and flexible. The examination with the latter one is less painful and is used widely. In this paper, we deal the examination with flexible cystoscope.

In flexible cystoscopy, images of the bladder are obtained from a camera embedded in the tip of flexible cystoscope and displayed in the monitor. However, there are three major difficulties for physicians to check the whole inner of the bladder completely. First, since the bladder has similar shape and color, sometimes beginner physicians lose track of the observation. Second, cystoscopic images are sometimes unclear due to halation. Third, it requires some experiments to control flexible cystoscope properly due to complex handling patterns of the equipment.

As a diagnostic support tool for beginner physicians in flexible cystoscopy, this paper presents a system for tracking the observation. To achieve the objective, it is considered to attach acceleration sensors or location sensors to a flexible cystoscope. In that case, it is necessary to obtain an approval for usage of the cystoscope according to the pharmaceutical low, as well as to buy the cystoscope which would be at least 10,000$. Another approach to track the observation in cystoscopy is mapping observed regions in a 3D space. As a representative 3D tracking system, Choi et al. proposed a robust segment-based object tracking system that uses the backside of a car image[3]. Their system measures depth information by calculating the enlargement factor of a target region. Ramisa et al. proposed to measure the distance between a single camera and a person[4]. However, since cystoscopic images are significantly unclear than the car image and human image, existing algorithms for 3D tracking[3, 4] could not be applied to cystoscopic images.

In this paper, we propose to discriminate the handling patterns of a flexible cystoscope. First, the proposed system extracts ZNCC-based optical flows[5] from cystoscopic images as features for estimating the handling patterns. Next, various features including the ZNCC-based optical flows are learned by a Random Forest classifier[6]. The classifier discriminates sequential handling patterns of the cystoscope by a time-series analysis. Finally, the observation in flexible cystoscopy is reproduced in a virtual 3D bladder we propose.

In section 2, we will introduce the process of flexible cystoscopy. In section 3, we will explain the proposed system. In section 4, we will examine the performance of the proposed system regarding the accuracy in estimating the handling patterns and tracking the observation in the virtual bladder.

2. FLEXIBLE CYSTOSCOPY

The human bladder is a hollow and balloon shaped organ that is broadly distinguished into seven regions; trigone, neck, left side wall, right side wall, posterior wall, dome, and anterior wall. Figure 1 shows images of the bladder. From the figure, we could perceive the images except neck are similar to each other. In addition, sometimes cystoscopic images are noisy due to halation which often occurs when the cystoscope is close to the bladder wall.

The flexible cystoscopy is conducted by a physician as below.

(1) The physician inserts a flexible cystoscope into a patient's urethra.

(2) The physician pushes the cystoscope slowly to the bladder.

(3) By adjusting the position of the cystoscope, the physician observes the whole inner of the bladder.

(4) The physician pulls the cystoscope after checking all the regions.

Figure 2 shows three handling patterns of the cystoscope to adjust the position of the cystoscope in step (3). The problem in this examination is that a beginner physician in the step (4) is sometimes unsure that all the regions were completely observed. Since oversight of severe legion would cause fatal case, diagnostic support tools for beginner physicians in flexible cystoscopy have been required.

3. PROPOSED SYSTEM

3.1 Overview

As mentioned in Sec. 2, the bladder can be distinguished into seven regions. However, the definitive region of each part is not defined. Considering a normal bladder, we construct a sphere bladder model which has seven regions. Figure 3 shows the virtual bladder we propose. And, Table 1 shows definitions of each region for the virtual bladder, which determined by an expert physician in flexible cystoscopy.

3.2 Preprocessing

Figure 4 shows the interface for a flexible cystoscopy. In the proposed system, first, ROI (Region of Interest) is set on the rectangle in the interface. The size of ROI is 300 pixel $\times$ 300 pixel. Next, Figure 5 (b) shows the cystoscopic image applied 8-bit gray scale transformation to Figure 5 (a). And,

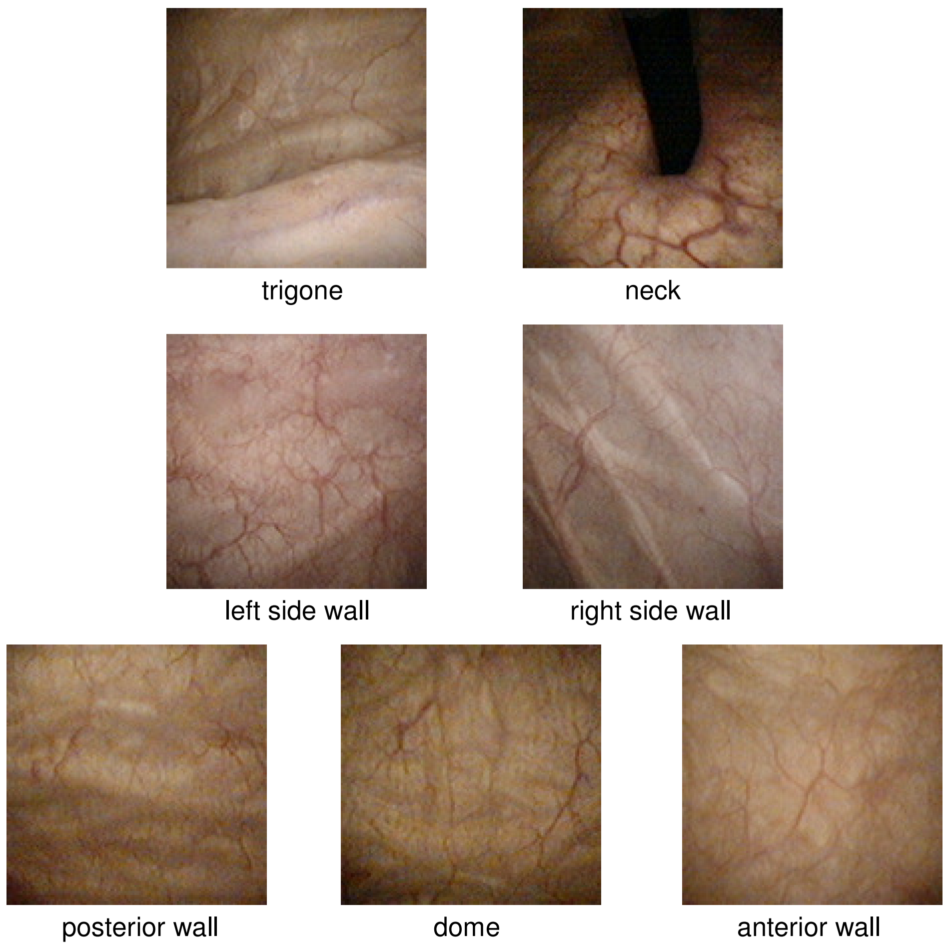


FIGURE 1: Example of Images Obtained from a Flexible Cystoscope.

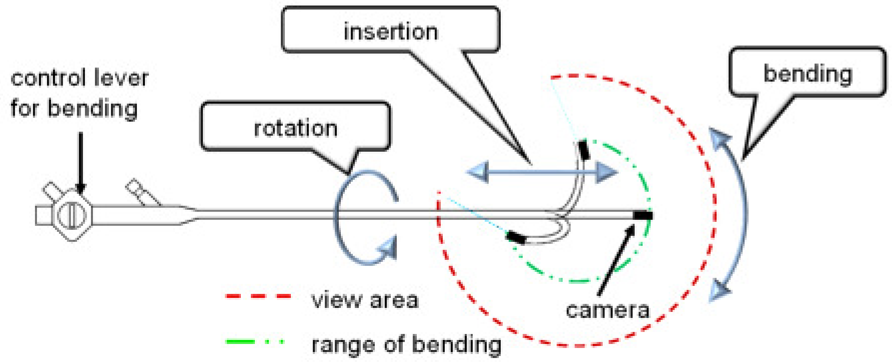


FIGURE 2: Three Handling Patterns of a Flexible Cystoscope.

Figure 5 (c) shows the image applied the histogram stretching to Figure 5 (b). Finally, Figure 5 (d) shows the image applied the selective local averaging to Figure 5 (c). Compared with Figure 5 (b), we could perceive that blood vessels are enhanced in the images of Figure 5 (d).

3.3 Extraction of Optical Flows

Approaches of well-known object tracking can be distinguished into gradient method and block matching method. Gradient method is effective on videos where target objects move slowly[7].

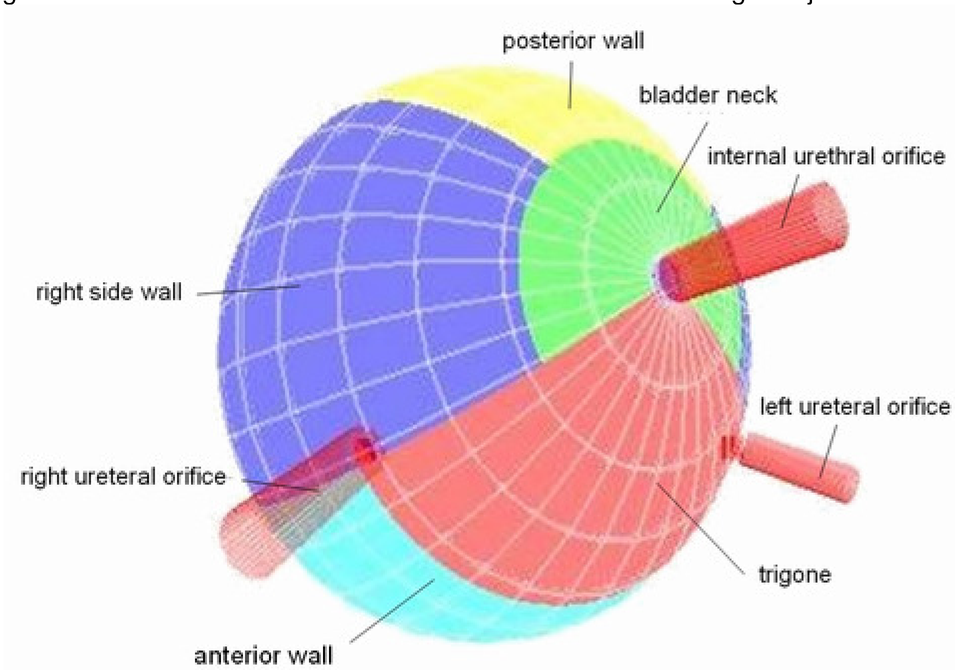


FIGURE 3: 3D Virtual Bladder We Propose.

Bladder region	**Definition**
Dome	More than lat. 70 degrees S, excluding the trigone defined below.
Bladder neck	More than lat. 70 degrees S, excluding the trigone defined below.
Trigone	Triangle part is surrounded with lat. 45 degrees S, a parallel of lines of longitude of long. 60 degrees E / long. 60 degrees W.
Posterior wall	The quadrangle surrounded with a parallel of lat. 60 degrees N/ lat. 70 degrees S, a line of longitude of long. 60 degrees E/ long. 60 degrees W.
Anterior wall	The quadrangle surrounded with a parallel of lat. 60 degrees N/ lat. 70 degrees S, a line of longitude of long. 120 degrees E/ long. 120 degrees W.
Right side wall	The reminded region in the east side.
Left side wall	The reminded region in the west side.

TABLE 1: Definition of the Virtual Bladder Map.

Meanwhile, block matching method is robust to quick movement of target objects, sudden illumination changes, and noises. Since cystoscopic images are often unclear due to noises such as halation and focus error, the proposed method applies the block matching[8]. As a robust method to measure the movements of a flexible cystoscope, the proposed method extracts ZNCC-based optical flows from consecutive cystoscopic images. In ZNCC (Zero-mean Normalized Cross-Correlation), an image is divided into $A \times B$ blocks and optical flows are extracted from each block by the following formula

$$R_{ZNCC} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(I(i,j)-\bar{I})(T(i,j)-\bar{T})}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}(I(i,j)-\bar{I})^2 \times \sum_{i=1}^{N}\sum_{j=1}^{M}(T(i,j)-\bar{T})^2}} \qquad (1)$$

where $I(i, j)$ is the pixel value at ith row and jth column in the template image, $T(i, j)$ is the pixel value at ith row and jth column in the target image, N is the search range towards x axis, and M is

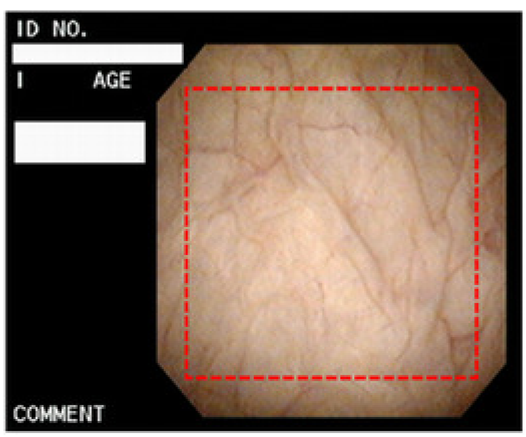
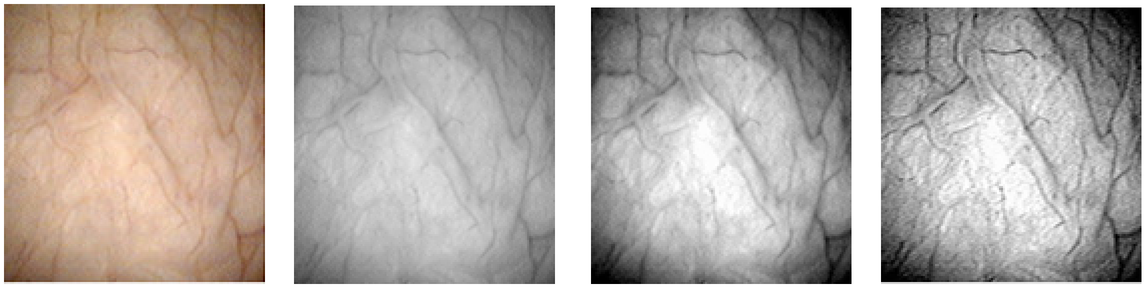


FIGURE 4: Interface for the Examination with a Flexible Cystoscope.



(a) (b) (c) (d)

FIGURE 5: Preprocessing for the Enhancement of Blood Vessels.

Handling pattern			
rotation	left	right	neutral
bending	up	down	neutral
insertion	push	pull	neutral

TABLE 2: Handling Patterns of Flexible Cystoscope.

the search range towards *y* axis. As an algorithm of ZNCC, the proposed system uses the one proposed by Yoo and Han[5].

3.4 Discrimination of Handling Patterns for a Flexible Cystoscope

Table 2 shows handling techniques of the flexible cystoscope. From the table, we can find that the combination of the handling techniques is up to 27 patterns. And, figure 6 shows an example of optical flows obtained by rotation (a) and insertion (b) when the cystoscope is 90 degrees down or zero degrees or 90 degrees up. From figure 6, we can find that the ZNCC-based optical flows are depended on the degree of rotation and insertion.

We define three features that improve the discrimination of the handling patterns as below

$$Rot_t = \sum_{i=1}^{t} R_i \tag{2}$$

$$Bend_t = \sum_{i=1}^{t} B_i \tag{3}$$

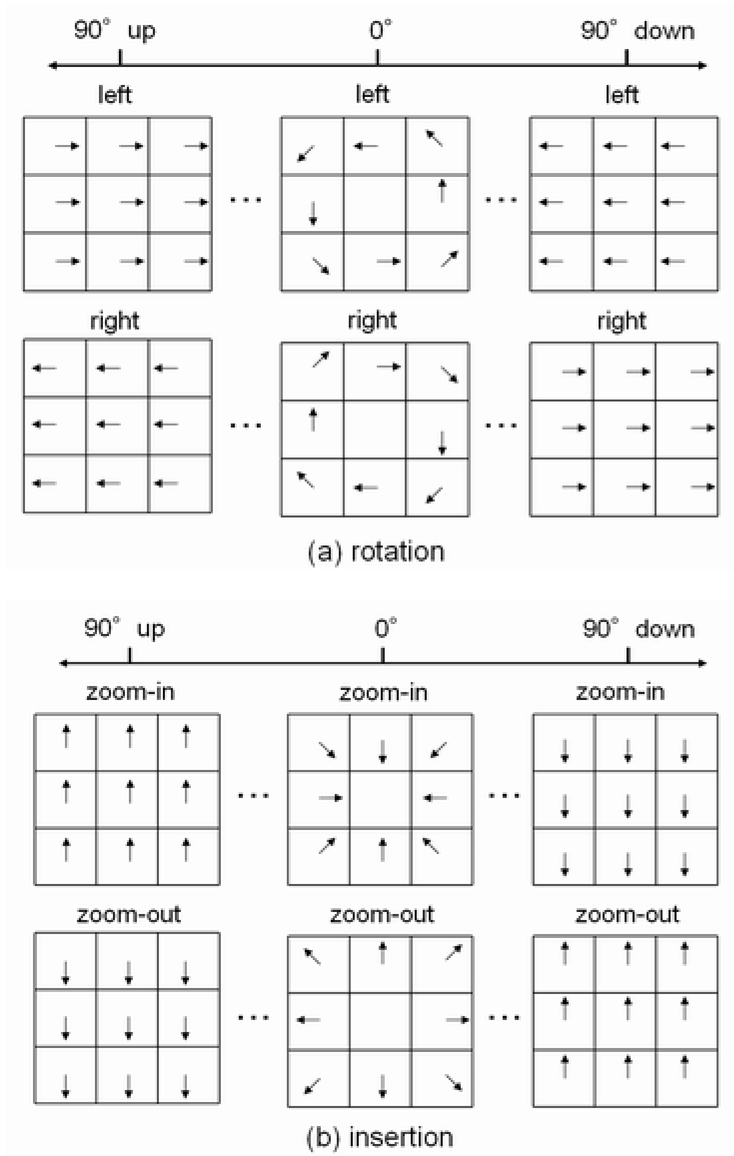


FIGURE 6: Example of Optical Flows Obtained by Rotation (a) and Insertion (b).

$$Ins_t = \sum_{i=1}^{t} I_i \tag{4}$$

where Rot_t is the degree of rotation at tth frame, R_i is the handling pattern of rotation at ith frame and returns 1 when R_i=left, -1 when R_i=right, and 0 when R_i=neutral. Similarly, $Bend_t$ is the degree of bending at tth frame, B_i is the handling pattern of bending at ith frame and returns 1 when B_i=up, -1 when B_i=down, and 0 when B_i=neutral, and Ins_t is the degree of insertion at tth frame, I_i is the handling pattern of insertion at ith frame and returns 1 when I_i=push, -1 when I_i=pull, and 0 when I_i=neutral.

Thus, the proposed system extracts features as $A \times B$ optical flows, Rot_t, $Bend_t$, and Ins_t from a cystoscopic image. By feeding all the features, a RF (random forest) classifier discriminates the 27 handling patterns frame by frame. RF is a noise-robust classification algorithm proposed by Breiman.

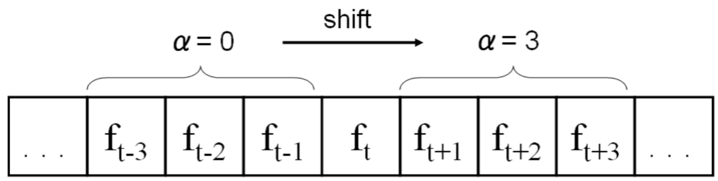


FIGURE 7: Selection of consecutive k frames whose prediction probabilities are similar to those of f_t (k=5).

3.5 Discrimination of Sequential Handling Patterns for a Flexible Cystoscope

In flexible cystoscopy, the same handling pattern would last for several frames. Hence, in a case that the discriminate handling for a frame f_t is different from the one for f_{t-1} and f_{t+1}, f_t is expected to be another handling pattern. Considering the case, we propose to correct the handling pattern for f_t. Figure 7 shows a process to discriminate sequential k frames, where k = 5 is configured in this case. The sequential k frames are determined by the following steps.

(1) Select the sequential k frames from $f_{t-k+\alpha}$ to $f_{t-1+\alpha}$ (Initially, $\alpha = 0$).

(2) Calculate a similarity of the class prediction probabilities for each frame.

(3) $\alpha = \alpha + 1$ if $\alpha < k$ and go back to step (1).

(4) Determine the k frames when the similarity in step (2) is maximum.

And then, the similarity Sim_t is calculated by the following formula

$$Sim_t[\alpha] = \sum_{j=1}^{27} | aprob_t[j] - prob_t[j] | \tag{5}$$

where $prob_t[j]$ is the class prediction probability of f_t for jth handling pattern, $aprob_t[j]$ is the average class prediction probability of selected k frames for jth handling pattern. Thus, the proposed method selects k frames whose average class prediction probability are similar to that of f_t. And then, the discriminated handling pattern for f_t is replaced by the jth handling pattern where $aprob_t[j]$ becomes maximal in k frames.

Next, regarding f_{num} as the number of replaced handling patterns (f_{num}=1,2,3...,k-1), f_{num} is increased in 1 by 1 (f_{num} is 1 in the initial step above). Then, regarding k as k', k' is set as k-f_{num}+1. Note that the replacement of k' frames is conducted when the following term is fulfilled.

$$P_B \quad > \quad P_A \tag{6}$$

where P_B is the average class prediction probability for a handling pattern B in k' frames and P_A is the average class prediction probability for a handling pattern A in k' frames. These terms can prevent a case that the handling pattern of the correct sequential k' frames is replaced to incorrect one. The method above can be applied to frames before f_t when the discrimination of the handling pattern for t+k frames is finished.

4. PROPOSED SYSTEM

4.1 Experimental Environment

We applied the proposed system to ten videos of flexible cystoscopy at the Kanazawa University Hospital in Japan. All the examinations were conducted by an expect physician. The flexible cystoscope used in the examinations is OLYMPUS CYF TYPE VA2. Regarding the videos, the format is AVI (24-bit color), the frame rate is 29.97 fps, and the average length is 115 seconds. Each of the videos was cut manually as they start from the scene that the image of the bladder wall appears at the first time to the scene that the image of the urethra appears at the first time after the start scene. T evaluate the performance of the proposed system, by observing the videos, the expert physician judged the handling patterns frame by frame. Then, a RF (Random Forest) classifier discriminated the handling patterns. The accuracy of estimating the handling patterns is evaluated by 10-fold cross validation[9] as below.

(1) Choose the first 1000 frames in each of the ten videos.

(2) Extract ZNCC-based optical flows and Rot_t, $Bend_t$, and Ins_t (refer to Sec. 3.4) from each of the 1000 frames.

Handling pattern	Number of frames	*Base line*	proposed method
neutral	877	70.6%	93.2%
left	1049	81.0%	94.5%
right	932	82.1%	93.4%
up	741	83.3%	95.0%
down	1214	82.0%	96.2%
push	801	76.0%	94.6%
right	932	82.1%	93.4%

TABLE 3: Average correct rate for each of the handling patterns. *base line* is the case that only ZNCC-based optical flows were learned by the Random Forest classifier.

No.	1	2	3	4	5
Cr	54.9%	81.9%	91.6%	91.2%	62.5%
Fr	6.4%	2.6%	1.2%	1.3%	5.4%
No.	**6**	**7**	**8**	**9**	**10**
Cr	83.5%	84.1%	82.9%	92.1%	90.4%
Fr	2.4%	2.3%	2.4%	1.1%	1.4%

TABLE 4: Parameter *Cr* and *Fr* obtained in the Experiment.

(3) Choose the 1000 frames in one of the videos as test data and the other 9000 frames as training data.

(4) RF learns the training data.

(5) RF discriminates the handling patterns frame by frame in the test data.

(6) Sequential handling patterns of the cystoscope are discriminated by the time series analysis described in Sec. 3.5.

(7) Repeat the procedure from (3) until all the 1000 frames are selected as test data.

4.2 Results

First, we discriminated the handling patterns of the flexible cystoscope in each of the 1000 frames. Table 3 shows the average correct rate in the discrimination. In this experiment, parameters for the Random Forest classifier were optimized, namely the number of the trees was configured as 525. As the table shows, the proposed method outperforms *base line*.

Next, we reproduced each of the cystoscopic examinations in the virtual bladder defined in Sec. 3.1. In each examination, observed regions were painted on the virtual bladder. To evaluate the performance of reproducing the observation, we assume that the physician could not observe one of the whole regions. Such a region is called as target region in the rest of this paper. For example, suppose that the physician could not observe trigone, the region of trigone in the virtual bladder is unpainted ideally and the other regions are painted. Here, we define *Cr* as the percentage of the unpainted area in the target region and *Fr* as the percentage of the painted area in the other region. Therefore, the ideal *Cr* is 100% and the ideal *Fr* is 0%. Table 4 shows the average *Cr* and *Fr* for each of the video. From the table, we can find that *Cr* has been more than 80% except the video No.1 and No.5.

4.3 Failure Cases

In Table 4, both of *Cr* and *Fr* for the video No.1 has been the worst among all the videos. Observing the video, we could find that bladder stones were floating around trigone. Figure 8 shows the stones marked in circle. In 114 frames, although the camera was being stopped, the block-matching method detected movement of the stones and the proposed system incorrectly judged the handling patterns in the 114 frames as left or insertion. The bladder stones are sometimes observed in cystoscopy. Therefore, we need to detect the white stones as noise. In Figure 8, average density of the stones in areas of each circle has been 223.4 (standard deviation is 15.9) while average density of all the 64072 frames in ten videos is 124.9. The densities above were

FIGURE 8: Bladder Stones Appeared Around Trigone in the Video No.1.

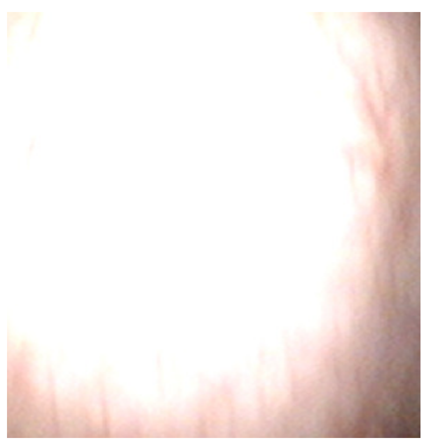


FIGURE 9: One of the Images where Halation Observed in the Video No.5.

measured from the original images obtained from the cystoscope. Hence, areas of the bladder stones in each image would be extracted using the density distributions of the whole image.

In Table 4, *Cr* and *Fr* for the video No.5 has been the second worst among all the videos. Observing the video No.5, we could find that halation lasted for 125 consecutive frames. Figure 9 shows an image of the halation. Since the halation covered overall area in each image, the block-matching method could not extract movement of the flexible cystoscope. Hence, in the case when handling patterns are judged incorrectly due to the halation, it is necessary to discriminate the handling patterns from handling patterns in other frames.

4.4 Future Works

From the experimental results shown in Sec. 4.2, this paper has indicated that our proposed system can be used as a trainer for beginner physicians. Although cystoscopic images are significantly unclear compared with images used in related works [3, 4, 10], it is seen that the proposed system works well on tracking the observation in a flexible cystoscopy. However, sometimes the tracking would fail due to the failure cases described in Sec. 4.3. In such case, it is necessary to estimate the actual position according to landmarks placed in the virtual bladder. One of the ways for generating landmarks is the use of HOG (Histogram of Oriented Gradients) which is robust against rotation and size difference. In addition, we would like to correct turbulent flows obtained from noisy images.

5. CONCLUSION

This paper has presented a system of tracking the observation in flexible cystoscopy. The proposed system discriminates three handling patterns of flexible cystoscope. To achieve this objective accurately, we proposed to extract ZNCC-based optical flows and three features that nd the represent the degree of the handling patterns from cystoscopic images. In addition, we proposed to discriminate sequential handling patterns of the flexible cystoscope. Experimental results using ten videos have shown the average correct ratio of the three handling patterns has been at least 90%. We also reproduced the observation in a flexible cystoscopy in a virtual 3D bladder we constructed.

Considering the failure cases for tracking the observation, we need to estimate the actual position in case of the tracking failure and correct turbulent flows obtained from cystoscopic images where bladder stones are floating and halation is happened. Besides, we would like to estimate the shape and size of the bladder using CT or MRI images.

6. REFERENCES

[1] M. Froehner, M. A. Brausi, H. W. Herr, G. Muto, U E. Studer. “Complications following radical cystectomy for bladder cancer in the elderly” European Urology, vol.56, no.3, pp.443-454, 2009.

[2] J. Key, D. Dhawan, D. K. Knapp, K. Kim, I. C. Kwon, K. Choi, J. F. Leary. “Method and apparatus for estimating the velocity vector of multiple vehicles on non-level and curved roads using a single camera” in Proc. SPIE 8225, Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues X, 82251F, 2012, pp.1-8.

[3] H C. Choi, S. Y Oh. “Robust segment-based object tracking using generalized hyperplane approximation” Pattern Recognition, vol.45, no.8, pp.2980-2991, 2012.

[4] A. Ramisa, G. Alenya, C. Torras. “Single image 3D human pose estimation from noisy observations” in Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp.2673-2680.

[5] J. C. Yoo and T. Han. “Fast normalized cross-correlation” Circ. Syst. Signal Process, vol.28, no.6, pp.819-843, 2009.

[6] L. Breiman. “Random Forests” Machine Learning, vol.45, no.1, pp.5-32, 2001.

[7] S.L. Tanimoto. “Template Matching in Pyramids” Computer Graphics and Image Processing, vol.16, no.4, pp.356-369, 1981.

[8] B.K.P Horn and B.G. Schunck. “Determining optical flow” Artif. Intell., vol.17, pp.185-203, 1981.

[9] M. Fosteller. “A *k*-sample slippage test for an extreme population” Annals of Mathematical Statistics, vol.19, no.1, pp.58-65, 1948.

[10] L. Beyang, S. Gould, and D. Koller. “Single image depth estimation from predicted semantic labels” in Proc. 2010 IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp.1253-1260.

[11] N. Dalal and B. Triggs. “Histograms of Oriented Gradients for Human Detection” in Proc.2005 IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp.886-893.