ENDOSENSORFUSION: PARTICLE FILTERING-BASED MULTI-SENSORY DATA FUSION WITH SWITCHING STATE-SPACE MODEL FOR ENDOSCOPIC CAPSULE ROBOTS USING RECURRENT NEURAL NETWORK KINEMATICS

by

Yasin Almalıoğlu

B.S., Computer Engineering, Boğaziçi University, 2015

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering
Boğaziçi University
2017

ACKNOWLEDGEMENTS

I would like to thank all the people who contributed in some way to the work described in this thesis. First, I thank my M.S. advisor Assoc. Prof. Taylan Cemgil for the support of my study and related research, for his patience and academic motivations he has provided to me. I have learned a lot from him.

I would also like to thank The Physical Intelligence Department Max Planck Institute Stuttgart, especially Mehmet Turan, for the collaboration in my research and their precious support.

Last but foremost, a very special gratitude goes to my wife, the true friend in my life, and to my family whose continuous support has always been with me.

ABSTRACT

ENDOSENSORFUSION: PARTICLE FILTERING-BASED MULTI-SENSORY DATA FUSION WITH SWITCHING STATE-SPACE MODEL FOR ENDOSCOPIC CAPSULE ROBOTS USING RECURRENT NEURAL NETWORK KINEMATICS

Ingestible wireless capsule endoscopy is an emerging minimally invasive diagnostic technology for inspection of the gastrointestinal (GI) tract and diagnosis of a wide range of diseases and pathologies. Medical device companies and many research groups have recently made substantial progresses in converting passive capsule endoscopes to active capsule robots, enabling more accurate, precise, and intuitive detection of the location and size of the diseased areas. A reliable, real time multi-sensor fusion functionality is crucial for localization of actively controlled next-generation endoscopic capsule robots. In this study, we propose a novel multi-sensor fusion approach based on switching observations model using non-linear kinematics learned by recurrent neural networks for real-time endoscopic capsule robot localization. Our method concerns the sequential estimation of a hidden state vector from noisy pose observations delivered by multiple sensors, a 5 degree-of-freedom (5-DoF) absolute pose measurement by a magnetic localization system and a 6-DoF relative pose measurement by visual odometry. For the inference of the model, Sequential Monte Carlo (SMC) methods known as particle filters are employed, which are effective for on-line inference. In addition, the proposed method is capable of detecting and handling sensor failures by ignoring corrupted data, providing the robustness of a medical device. Detailed analyses and evaluations made using ex-vivo experiments on a porcine stomach model prove that our system achieves high translational and rotational accuracies for different types of endoscopic capsule robot trajectories.

ÖZET

KAPSÜL ENDOSKOPİ ROBOTLARI İÇİN DEĞİŞEN DURUM-UZAY MODELİ İLE YİNELENEN YAPAY SİNİR AĞLARI KULLANARAK PARÇACIK FİLTRELEME TEMELLİ ÇOKLU DUYARGA VERİSİ İLİŞKİLENDİRMESİ

Yutulabilir kablosuz kapsül endoskopi, mide-bağırsak kanalı incelemeleri ve pek çok hastalık ve patolojinin tanısı için kullanılan, gelişmeye açık, tahriş miktarı düşük bir tanılama teknolojisidir. Tıbbi cihaz şirketleri ve pek çok araştırma grubu pasif kapsül endoskopinin daha doğru, hassas ve hastalıklı bölgelerin genişliğini ve yerini sezgisel olarak tespit edebilecek şekilde geliştirmek için önemli ilerlemeler kaydetti. Aktif olarak kontrol edilen, yeni-nesil kapsül endoskopi robotlarının konumlandırılması için güvenilir, gerçek-zamanlı çoklu duyarga ilişkilendirme özelliği hayati önem taşımaktadır. Bu çalışmada, kapsül endoskopi robotunun gerçek-zamanlı konumlandırılması için yinelenen yapay sinir ağları ile modellenen doğrusal olmayan robot kinematiği kullanılarak, değişen gözlem modeli temelli yeni bir, çoklu duyarga ilişkilendirme yaklaşımı önerildi. Sunulan yöntem çoklu duyargalardan, manyetik konumlandırma sistemi ile elde edilen 5 serbestlik-dereceli mutlak poz ölçümü ve görsel odometri yöntemi ile elde edilen 6 serbestlik-dereceli göreceli poz ölçümü, gelen gürültülü verilerden gizli durum vektörü için ardışık olarak kestirim yapılmasıyla ilgilenmektedir. Model üzerinde kestirim yapabilmek için, çevrimiçi kestirim için etkili olan, parçacık filtresi olarak da bilinen ardışıl Monte Carlo yöntemleri kullanılmıştır. Ex-vivo deneyler kullanılarak domuz midesi modelinde yapılan detaylı analizler ve hesaplamalar, sistemin farklı türde kapsül endoskopi robotunun gezingeleri için yüksek ilerleme ve dönme doğruluklarına sahip olduğunu ispatlıyor.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . iii
ABSTRACT . . . iv
ÖZET . . . v
LIST OF FIGURES . . . viii
LIST OF TABLES . . . x
LIST OF SYMBOLS . . . xi
LIST OF ACRONYMS/ABBREVIATIONS . . . xii
1. INTRODUCTION . . . 1
1.1. Bayesian Estimation . . . 3
1.1.1. State-Space Models . . . 3
1.1.1.1. Switching State-Space Models . . . 5
1.1.2. Kalman Filter . . . 6
1.1.2.1. Extended Kalman Filter . . . 7
1.1.2.2. Unscented Kalman Filter . . . 8
1.1.3. Sequential Monte Carlo . . . 9
1.1.3.1. Sequential Importance Re-sampling . . . 11
1.2. Experimental Setup . . . 14
1.2.1. Magnetically Actuated Soft Capsule Endoscopes (MASCE) . . . 14
1.2.2. Magnetic Localization System . . . 14
1.2.3. Monocular Visual Odometry . . . 15
1.2.4. Hand-eye Calibration . . . 19
2. MULTI-SENSOR FUSION WITH SWITCHING STATE-SPACE MODEL . 20
2.1. The Sequential Bayesian Model and Problem Statement . . . 20
2.2. Proposal Distributions . . . 22
2.3. The Particle Filter Algorithm . . . 24
2.4. RNN-based Kinematics Model . . . 25
2.5. Dataset . . . 30
3. EXPERIMENTS AND RESULTS . . . 32
4. CONCLUSIONS . . . 39

REFERENCES 41

LIST OF FIGURES

Figure 1.1. Graphical model of a state space model. The latent and observation processes are denoted by single and double circles, respectively. . . 4

Figure 1.2. Kalman Filter Algorithm 7

Figure 1.3. Extended Kalman Filter Algorithm 8

Figure 1.4. Unscented Kalman Filter Algorithm 10

Figure 1.5. SIR Filter Algorithm 12

Figure 1.6. Schematic drawing of the magnetic localization technique. A capsule robot is manipulated by an external magnetic coil array consisting of nine electromagnets [1]. 15

Figure 1.7. Actuation system of the MASCE [1]. 16

Figure 1.8. The transformations between different frames at pose i and pose $i+1$ [2]. 18

Figure 2.1. The overall switching state-space model. Observable variables and hyper-parameters are denoted by double circles and gray circles, respectively. 21

Figure 2.2. Example ARS sampling result for $\log(\sigma_{k,t})$. The piecewise hull and the generated samples are shown. 24

Figure 2.3. Initialization of the particle filter algorithm used in the multi-sensor fusion model. 25

Figure 2.4. Sequential updates of the particle filter algorithm used in the multi-sensor fusion model. 26

Figure 2.5. Re-sampling method of the particle filter algorithm used in the multi-sensor fusion model. 27

Figure 2.6. Data flow through the hidden units of the LSTM [3]. 27

Figure 2.7. Experimental setup [4]. 29

Figure 2.8. Sample frames from the dataset used in the experiments [4]. . . . 31

Figure 3.1. Sample trajectories comparing the multi-sensor fusion result with ground truth and sensor data. 33

Figure 3.2. Posterior probability of $\mathbf{s}_{k,t}$ parameter for endoscopic RGB camera (top) and for magnetic localization system (bottom). The switch parameter, $\mathbf{s}_{k,t}$, reflects the failure times accurately. 34

Figure 3.3. The minimum mean square error (MMSE) of $\alpha_{k,t}$ for endoscopic RGB camera (top) and for magnetic localization system (bottom). 35

Figure 3.4. Evolution of the $\sigma^{\alpha}_{k,t}$ parameter for the sensors. $\sigma^{\alpha}_{k,t}$ does not tend to increase during sensor failure periods. 36

Figure 3.5. Translational (top) and rotational (bottom) RMSEs for multi-sensor fusion, visual localization and magnetic localization. 37

LIST OF TABLES

Table 1.1. Endoscopic camera specifications used for the experiments. 17

LIST OF SYMBOLS

d_k	The number of possible observation models
$f(\cdot)$	Non-linear state transition function
$\mathbf{F}$	State transition matrix
$\mathbf{G}$	Observation matrix
$h_{k,s_{k,t},t}(\cdot)$	The non-linear observation function
n	The number of sensors
$q(\cdot)$	Proposal distribution
$\mathbf{Q}$	Process noise covariance
$\mathbf{R}$	Observation noise covariance
$\mathbb{R}^m$	m-dimensional Euclidean space
$s_{k,t}$	The switch variable
t	Index of time sequences
$\mathbf{v_t}$	White noise
$\mathbf{w}_{k,s_{k,t},t}$	The observation noise
$\mathbf{x}_t$	Hidden states of the model
$\mathbf{z}_{k,t}$	Multi-sensory observations
$\alpha_{k,j,t}$	The prior probability for the switch parameter $s_{k,t}$
$\delta(\cdot)$	The Dirac delta function
$\mu(\cdot)$	Prior distribution

LIST OF ACRONYMS/ABBREVIATIONS

ARS	Adaptive Rejection Sampling
DoF	Degree-of-freedom
EKF	Extended Kalman Filter
ESS	Effective Sample Size
GI	Gastrointestinal
LSTM	Long Short-Term Memory
MASCE	Magnetically actuated soft capsule endoscope
MCMC	Markov Chain Monte Carlo
MMSE	The minimum mean square error
OF	Optical flow
pdf	Probability density function
RNN	Recurrent Neural Network
SSM	State-Space Model
SSSM	Switching State-Space Model
SVD	Singular Value Decomposition
UKF	Unscented Kalman filter

1. INTRODUCTION

Following the advances in material science in last decades, milli-scale endoscopic capsule robots with an on-board camera and wireless image transmission device have been commercialized and used in hospitals (FDA approved) since 2001. Untethered pill-size, swallowable capsule endoscopes are used to access to regions of the gastrointestinal (GI) tract that were impossible to access before, to diagnose diseases such as the inflammatory bowel disease, the ulcerative colitis and the colorectal cancer, and has reduced the discomfort and sedation related work loss issues [2, 5–8]. However, the capsule endoscopy is limited to passive monitoring of the GI-tract via optical imaging as clinicians do not have active control over the capsule's position, orientation, and functions. The control over capsule's position, orientation, and functions would give a doctor more precise reachability of targeted body parts and more intuitive and correct diagnosis opportunity [4, 9–13]. Several groups have recently proposed active, remotely controllable robotic capsule endoscope prototypes equipped with additional functionalities such as localized drug delivery, biopsy and other medical functions [1, 10, 14–22]. These medical functions require an active motion control with a reliable and precise real time pose estimation. In the last decade, many different approaches were developed for real time endoscopic capsule robot localization including received signal strength (RSS), time of flight and time difference of arrival (ToF and TDoA), angel of arrival (AoA) and RF identification (RFID)-based methods.

The advantage of electromagnetic wave-based techniques is that there is no need for an additional equipment apart from wireless biomedical sensors and the localization is not affected by actuating magnetic field unlike DC magnetic field strength-based techniques. The disadvantage of these techniques is that high-frequency electromagnetic waves attenuate quicker than magnetic waves while low-frequency electromagnetic waves provide a low precision. On the other hand, the advantage of magnetic field strength-based techniques is that actuation system and localization system can be both executed based on magnetic forces and torques. The disadvantage of magnetic sensor based localization is interference from environment into the magnetic fields.

Hybrid techniques based on the combination of different sources such as magnetic sensors, radio frequency (RF) sensors and RGB camera sensors are more beneficial for obtaining more reliable and relevant localization if the fusion strategy eliminates weaknesses of both sensors and elevates strengths of them. Sensor fusion techniques applied for endoscopic capsule robots include fusion of RF electromagnetic signal and video, fusion of RF electromagnetic signals and magnetic sensors data and fusion of magnetic sensors data and video. First subgroup of hybrid techniques fuses RF signals and video for localization of the capsule robot [23, 24]. [24] asserts that using both RF signal and video as input, achieves millimetre accuracy while previous techniques achieves a few centimetres accuracy. One drawback of these studies is that most techniques for data fusion have been based on Kalman filtering, which works best for linear systems, while the kinematics and dynamics of capsule robots are nonlinear in orientation. In the second group, RF signal and magnetic data are fused for the localization of the capsule robot [24–26]. In these studies, a localization method that has high accuracy for simultaneous position and orientation estimation, has been investigated. In the third group of hybrid techniques, video and magnetic data are fused for the localization of the capsule robot [27]. An ultrasound imaging-based localization combined with magnetic field-based localization is introduced by [27].

Although some of these state-of-the-art sensor fusion techniques have achieved remarkable accuracy for the tracking and localization task of a capsule robot, they are not able to detect and autonomously handle sensor faults, and additionally several techniques using RF localization require complex signal corrections to account for attenuation and propagation of RF signals inside human body tissues. In addition, most previous models use relatively simple dynamic models for the capsule, whereas performance would be greatly improved by a more accurate model of the system. Lastly, previously demonstrated methods generate inaccurate estimations in cases where noise from the environment and the actuation system interferes with one or more components of the localization system.

In this thesis, we propose a novel multi-sensor fusion algorithm for capsule robots based on switching state space models with particle filtering using the endoscopic

capsule robot dynamics modelled by Recurrent Neural Networks (RNNs), which can handle sensor faults and non-linear motion models. The main contributions of our thesis are as follows:

- To the best of our knowledge, this is the first multi-sensor data fusion approach that combines a switching observation model, a particle filter approach, and a recurrent neural network developed for the endoscopic capsule robot and hand-held endoscope localization.
- We propose a sensor failure detection system for endoscopic capsule robots based on probabilistic graphical models with efficient proposal distributions applied onto the particle filtering. The approach can be generalized to any number of sensors and any mobile robotic platforms.
- No manual formulation is required to determine a probability density function that describes the motion dynamics, contrary to traditional particle filter and Kalman filter based methods.

The thesis is constructed as follows. A review of the state-space models and Bayesian filtering methods is given in Section 1.1. Section 1.2 presents the dataset and the experimental setup. Section 2 introduces the proposed multi-sensory data fusion and the filtering algorithm in detail. Section 3 shows the experimental results for 6-DoF multi-sensor data fusion of the endoscopic capsule robot. Section 4 concludes with future directions.

1.1. Bayesian Estimation

1.1.1. State-Space Models

State space models (SSM) provide a framework to handle sequential data, which assumes a data sequence with processes $z_{1:T} \triangleq \{z_t\}_{t=1}^{T}$ and $x_{0:T} \triangleq \{x_t\}_{t=0}^{T}$. The model

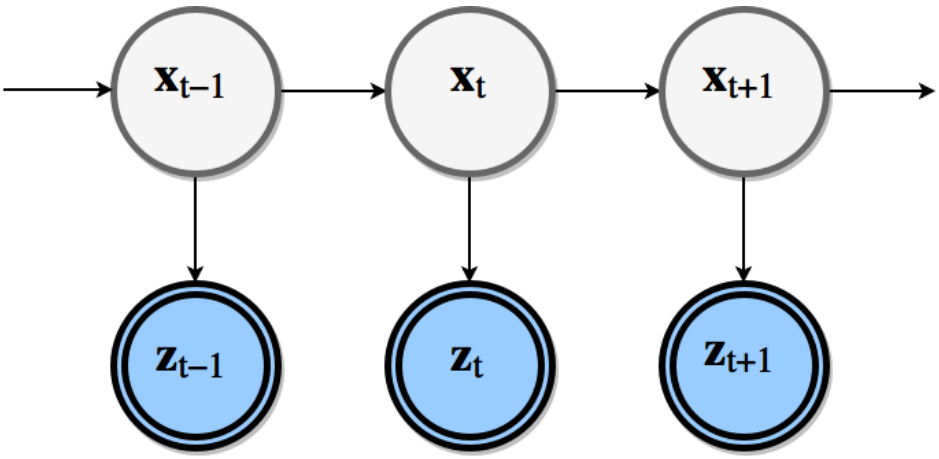


Figure 1.1. Graphical model of a state space model. The latent and observation processes are denoted by single and double circles, respectively.

is given on the compact form in terms of conditional distributions:

$$x_0 \sim \mu(x_0) \tag{1.1}$$

$$\mathbf{x}_{t+1} \mid \mathbf{x}_t \sim f(x_{t+1} \mid x_t) \tag{1.2}$$

$$\mathbf{z}_t \mid \mathbf{x}_t \sim h(z_t \mid x_t) \tag{1.3}$$

where $x_t \in X \in \mathbb{R}^n$ is the latent state, $z_t \in Z \in \mathbb{R}^m$ is the observation at time t, an initial latent state x_0 with a prior distribution $\mu(x_0)$, a state transition function $f(x_{t+1} \mid x_t)$ and an observation function $h(z_t \mid x_t)$.

Another description of sequential models like state-space models is graphical model [28,29]. Figure 1.1 shows the corresponding graphical representation of an SSM, where the latent state process and observed process are presented in single and double circles, respectively. In the graphical model, we see that the state x_t is conditionally independent of the states $x_{0:t-2}$ given the previous state x_{t-1} due to the Markov property of the model, meaning all the past information is contained in the state at time $t-1$. In addition, the observations are mutually independent given the latent states, as the arrows show the conditional dependency.

1.1.1.1. Switching State-Space Models. Switching state-space models (SSSM) are a class of time series models, which have novel applications in statistics, robotics, econometrics and signal processing. SSSMs are composed of:

- A switch variable with a statistical model, which adapts the behaviour of the statistical SSM that switches from one observation model to another
- A latent state vector that is evolved by a stochastic evolution function
- An observation vector that is related to the state vector according to an observation function.

The statistical model for the switch variable can be formulated in different ways such as:

- The switch parameter $\mathbf{s}_t$ is defined to be identically and independently distributed random variable:

$$Pr(\mathbf{s}_t \mid \mathbf{s}_{1:t-1}) = Pr(\mathbf{s}_t) \tag{1.4}$$

- The $\mathbf{s}_t$ is defined to be a homogeneous, discrete-time Markov chain a transition function:

$$Pr(\mathbf{s}_t \mid \mathbf{s}_{1:t-1}) = Pr(\mathbf{s}_t \mid \mathbf{s}_{t-1}) \tag{1.5}$$

Statistical structures for these Markov models have been widely studied in the literature, especially when the SSM is Gaussian and conditionally linear [30,31]. Multiple model algorithm and pseudo-Bayes algorithms based on Gaussian mixture approximations are defined to solve the estimation problem [32]. Several sequential Monte Carlo methods are used for approximate inference in linear and non-linear jump Markov systems that are applied in digital communications multiple target tracking, fault detection and image processing [33–35].

1.1.2. Kalman Filter

In a sequential estimation problem, we are mostly interested in the value of x_t given all the observations up to time t, $Pr(x_t \mid z_{1:t})$, which is called the *filtering* problem. The posterior distribution can be expressed recursively:

$$\begin{aligned} Pr(x_t \mid z_{1:t}) &\propto h(z_t \mid x_t) Pr(x_t \mid z_{1:t-1}) \\ &= h(z_t \mid x_t) \int f(x_t \mid x_{t-1}) Pr(x_{t-1} \mid z_{1:t-1}) d(x_{t-1}). \end{aligned} \tag{1.6}$$

Thus, the posterior density of the filtering is proportional to the multiplication of the likelihood $h(z_t \mid x_t)$ by the prediction density $Pr(x_t \mid z_{1:t-1})$, which is impossible to calculate analytically in most problems. However, the posterior density can be expressed in closed form when the state transition function $f(x_t \mid x_{t-1})$, the observation function $h(z_t \mid x_t)$ and the prior are Gaussian and linear. The SSM has the following form:

$$x_0 \sim \mathcal{N}(\mu_0, \Sigma_0) \tag{1.7}$$

$$\mathbf{x}_{t+1} \mid \mathbf{x}_t \sim \mathcal{N}(\mathbf{F}\mathbf{x}_t), \mathbf{Q}) \tag{1.8}$$

$$\mathbf{z}_t \mid \mathbf{x}_t \sim \mathcal{N}(\mathbf{G}\mathbf{x}_t), \mathbf{R}) \tag{1.9}$$

where $\mathbf{F}$ and $\mathbf{G}$ are state transition and observation matrices, the positive semi-definite matrices $\mathbf{Q}$ and $\mathbf{R}$ are process and observation noise covariances, respectively. The algorithm of the Kalman filter can be found in Figure 1.2.

The Kalman filter algorithm takes noisy measurements from sensors and estimates state variables more accurate than each measurement alone. The model and the measurements involved can be estimated up to an extend due to model approximations or noise respectively. Prediction of the system state is averaged with new measurements using weights. These weights decide which measurements are more accurate, which are computed from the covariance that is the estimated uncertainty of the state prediction.

Prediction

$\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1}$; A priori state estimate

$\mathbf{P}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t|t-1}\mathbf{F}_t^T + \mathbf{Q}_t$; A priori covariance estimate

Update

$\hat{\mathbf{y}}_t = \mathbf{z}_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}$; Measurement pre-fit residual

$\mathbf{S}_t = \mathbf{R}_t + \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T$; Innovation covariance

$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^T\mathbf{S}_t^{-1}$; Optimal Kalman gain

$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\hat{\mathbf{y}}_t$; A posteriori state estimate

$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}$; A posteriori covariance estimate

$\hat{\mathbf{y}}_{t|t} = \mathbf{z}_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t}$; Measurement post-fit residual

Figure 1.2. Kalman Filter Algorithm.

This process runs iteratively, resulting in a recursion with information only from the previous time step.

The Kalman filter can handle missing observation at time t, z_t, making use of all available information at time t. Additionally, The Kalman filter can make predictions by repeating the prediction step of the filter k more times to obtain $Pr(x_t + k \mid y_{1:t})$. However, the Kalman filter suffers from the dependency on the strict linearity and Gaussian form assumptions and, thus, is not applicable to to non-linear and non-Gaussian models.

1.1.2.1. Extended Kalman Filter. To avoid the linearity constrain of the Kalman filter, the Extended Kalman Filter (EKF) is used as an extension, which allows the transition and observation functions to be non-linear with Gaussian noises. The algorithm of the EKF is shown in Figure 1.3. EKF linearises the model with first-order Taylor

approximations with the following equations:

$$\mathbf{F}_{t-1} = \frac{\partial f}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{t-1|t-1}} \tag{1.10}$$

$$\mathbf{H}_{t} = \frac{\partial h}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{t|t-1}} \tag{1.11}$$

However, if the model is highly non-linear, the first-order approximations can lead to large errors, resulting in a divergence of the filter. Even though higher order extensions are described for the Kalman filter, these are shown to be effective only when the measurement noise is small [36].

Prediction

$\hat{\mathbf{x}}_{t|t-1} = f(\hat{\mathbf{x}}_{t-1|t-1})$; A priori state estimate

$\mathbf{P}_{t|t-1} = \mathbf{F}_{t-1}\mathbf{P}_{t-1|t-1}\mathbf{F}_{t-1}^{T} + \mathbf{Q}_{t-1}$; A priori covariance estimate

Update

$\hat{\mathbf{y}}_{t} = \mathbf{z}_{t} - h(\hat{\mathbf{x}}_{t|t-1})$; Measurement pre-fit residual

$\mathbf{S}_{t} = \mathbf{R}_{t} + \mathbf{H}_{t}\mathbf{P}_{t|t-1}\mathbf{H}_{t}^{T}$; Innovation covariance

$\mathbf{K}_{t} = \mathbf{P}_{t|t-1}\mathbf{H}_{t}^{T}\mathbf{S}_{t}^{-1}$; Optimal Kalman gain

$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_{t}\hat{\mathbf{y}}_{t}$; A posteriori state estimate

$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_{t}\mathbf{H}_{t})\mathbf{P}_{t|t-1}$; A posteriori covariance estimate

Figure 1.3. Extended Kalman Filter Algorithm.

1.1.2.2. Unscented Kalman Filter. The Unscented Kalman Filter (UKF) aims to improve on the EKF when the state and observation equations are highly non-linear, making use of a set of discretely sampled points to parametrise the mean and covariance [37]. The UKF replaces the first-order approximations made by the EKF with a set of carefully chosen sample points called sigma points. The UKF is based on the Unscented Transform that is used to sample sigma points. True mean and covariance of the states are captured by the sigma points. The UKF then propagates the sigma points through the non-linear system dynamics capturing the posteriori mean

and covariance with the third order accuracy for any non-linear modality [38]. The UKF eliminates the need for derivation of the Jacobian of transition and observation functions. The UKF algorithm is described in Figure 1.4.

In the prediction step, the filter computes the prior using the non-linear process model $f(\cdot)$. Sigma points $\mathcal{X}$ and their corresponding weights W^m, W^c are generated and the sigma points are passed through $f(\cdot)$, which projects the sigma points forward in time according to the process model, forming the new prior. The transformed sigma points are denoted as $\boldsymbol{\mathcal{Y}}$. The mean and covariance of the prior are computed using the unscented transform on the transformed sigma points:

$$\bar{\mathbf{x}}, \bar{\mathbf{P}} = UT(\mathcal{Y}, W^m, W^c, \mathbf{Q}) \tag{1.12}$$

.

The UKF performs the update in measurement space. The sigma points of the prior are converted into measurements using the measurement function $h(\cdot)$. The mean and covariance of these points using the unscented transform, which are denoted as $\mu_{\mathbf{z}}$ and $\mathbf{S}$, respectively.

These non-linear extensions all aim to modify the Kalman filter to allow non-linear and non-Gaussian models with certain approximations which is exactly the same model when it is linear and Gaussian. Therefore, they can fail for highly non-linear and non-Gaussian models due to ineffective estimates of the state and covariances.

1.1.3. Sequential Monte Carlo

We would like to extend the Kalman Filter to any transition and observation processes $f(\cdot)$ and $h(\cdot)$ without making linear and Gaussian approximations. Thus, these models are not Gaussian filtering densities and the mean and covariances are not sufficient to construct the pdf. A more thorough approach is achieved by targeting the exact pdf. Since the calculation of these densities results in intractable integrals in

Prediction

$\mathcal{Y} = f(\mathcal{X})$

$\hat{\mathbf{x}}_{t|t-1} = \sum w^m \mathcal{Y}$; A priori state estimate

$\mathbf{P}_{t|t-1} = \sum w^c (\mathcal{Y} - \hat{\mathbf{x}}_{t|t-1})(\mathcal{Y} - \hat{\mathbf{x}}_{t|t-1})^T + \mathbf{Q}_{t-1}$; A priori covariance estimate

$\mathcal{Z} = h(\mathcal{Y})$

$\mu_{\mathbf{z}} = \sum w^m \mathcal{Z}$

Update

$\hat{\mathbf{y}}_t = \mathbf{z}_t - \mu_{\mathbf{z}}$; Measurement pre-fit residual

$\mathbf{S}_t = \sum w^c (\mathcal{Z} - \mu_{\mathbf{z}})(\mathcal{Z} - \mu_{\mathbf{z}})^T + \mathbf{R}_{t-1}$

$\mathbf{K}_t = [\sum w^c (\mathcal{Y} - \hat{\mathbf{x}}_{t|t-1})(\mathcal{Z} - \mu_{\mathbf{z}})^T]\mathbf{S}_t^{-1}$; Optimal Kalman gain

$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \hat{\mathbf{y}}_t$; A posteriori state estimate

$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T$; A posteriori covariance estimate

Figure 1.4. Unscented Kalman Filter Algorithm.

general, we resort to Monte Carlo methods that have the potential of approximating the target densities up to a negligible errors with adjustable Monte Carlo sample size. Markov Chain Monte Carlo (MCMC) methods are widely used to draw approximate samples from complex distributions in a general Bayesian framework. The target probability density is often required in closed form up to a proportionality constant, which makes standard MCMC methods inapplicable to sample from $p(x_t \mid z_{1:t})$ with an unknown form. However, MCMC methods can be applied to draw approximate samples from the joint *smoothing* distribution:

$$p(x_{0:t} \mid z_{1:t}) \propto \pi(x_0) \prod_{s=1}^{t} f(x_s \mid x_{s-1}) h(z_s \mid x_s) \tag{1.13}$$

where the marginal distribution can be calculated to obtain the filter distribution.

While the Kalman filter can be extended to arbitrary transition and observation processes thanks to Equation 1.13 in principle, it is inappropriate for many cases especially sequential observations. In a sequential estimation problem, $p(x_t \mid z_{1:t})$ is often

the objective density with a streaming data but it is not trivial to recursively update the samples with MCMC alone. Alternative Monte Carlo methods such as sequential importance sampling are investigated to approximate the filter densities [39]. However, the major problem of these methods is that all but one of the importance weights approaches to zero with time, called *degeneracy* problem, and it becomes necessary to recreate the samples by restarting MCMC from scratch. The whole trajectory $x_{0:t}$ must be sampled again and the complexity increases with time, making MCMC impractical for sequential inference.

1.1.3.1. Sequential Importance Re-sampling. Sequential Monte Carlo methods, also known as *particle filters*, have been proposed to overcome the restrictions of the Kalman Filter. Rather than approximating the objective filter probability distributions as Gaussian, a set of possible draws, called *particles*, are used to represent the intractable filter densities. These *particles* are sequentially updated according to the fitness of arriving observations. The resulting method is known as the bootstrap or the Sampling Importance Re-sampling (SIR) filter [40]. All the densities involving the latent state x_t are represented by the discrete distribution composed of Monte Carlo samples $\{x_t^{(i)}\}_{i=1}^N$ with the probabilities:

$$p(x_t) \simeq \frac{1}{N} \sum_{i=1}^{N} \delta(x_t - x_t^{(i)}) \tag{1.14}$$

where $\delta(\cdot)$ is the Dirac delta function. The integrals with respect to this discrete density are then approximated by the following equation:

$$\mathbf{E}(g(\mathcal{X})) = \int g(x)p(x)dx \simeq \frac{1}{N} \sum_{i=1}^{N} g(x - x^{(i)}) \tag{1.15}$$

which becomes exact as $N \to \infty$ by the weak law of large numbers.

The one-step predictive distribution $p(x_t|z_{1:t-1})$ is approximated by samples $\{\hat{x}_t^{(i)}\}$ that are generated from the state transition density $f(x_t \mid x_{t-1}^i)$. New observation z_t is used to weight the particles with the likelihood density $h(y_t \mid \hat{x}_t^i)$, which is ignored if the

observation is missing. New predictive particles are sampled according to probabilities proportional to the particle weights. Finally, the set of re-sampled draws $\{x_t^{(i)}\}$ are used to approximate the filter density $p(x_t \mid z_{1:t})$. The algorithm for the SIR filter is shown in Figure 1.5.

Initialisation from prior: $\{x_0^{(i)}\} \backsim \pi(x_0)$

for $t = 1$ to T **do**

Prediction: Sample each particle $\hat{x}_t^i \backsim f(\cdot \mid x_{t-1}^{(i)})$

Weight particles: Assign the weights to each particle and normalise them:

$$w_t^{(i)} \propto h(z_t \mid \hat{x}_t^{(i)}) \tag{1.16}$$

Re-sample particles: Draw N samples of x_t with support $\{\hat{x}_t^{(i)}\}$ and probability masses $\{w_t^{(i)}\}$

end for

Figure 1.5. SIR Filter Algorithm.

The SIR filter has fewer restrictions than the Kalman filter and it requires:

- Sampling from the prior $\pi(x_0)$,
- Sampling from the state transition distribution $f(x_t|x_{t-1})$,
- The likelihood distribution $h(z_t|x_t)$ is known up to a proportionality constant,

which can be pre-defined functions or estimated with a supervised or unsupervised learning methods [41, 42].

The filter algorithm is simple but it can perform poorly when there is a significant difference between the regions of the state-space in the filtering density $p(x_t|z_{1:t-1})$ and the likelihood $h(z_t|x_t)$, resulting in a set of uneven weights and few effective predictive particles $\hat{x}_t^{(i)}$.

Unlike the SIR filter algorithm re-samples the particles to create a new set of unweighed particles, sequential MC methods that outputs weighted particles are proposed. These MC methods calculates the discrete approximated density with particles $x_t^{(i)}$ that has weight $w_t^{(i)}$ becoming the probability mass of each particle. The integrals calculated with respect to the density is approximated as:

$$\mathbf{E}(g(\mathcal{X})) = \int g(x)p(x)dx \simeq \sum_{i=1}^{N} g(x^{(i)})w^{(i)} \tag{1.17}$$

In the state space model, the sequential imputation can employed by sequentially sampling $x_t^{(i)} \sim p(x_t \mid x_{t-1}^{(i)}, z_t)$ with the normalised weights $w_t^{(i)} \propto p(z_t \mid x_{t-1}^{(i)})w_{t-1}^{(i)}$. Instead of re-sampling the particles at each time step, the weights are sequentially updated, which requires $p(x_t \mid x_{t-1}, z_t)$ and $p(z_t \mid x_{t-1})$ to be known.

The Sequential Importance Sampling (SIS), a more generic method, allows an arbitrary proposal distribution to be sampled from and accounted for by the importance weight [43]. The SIS filter does not re-sample the particles propagated through the state equation. In the re-sampling, the weights become increasingly uneven and the mass of a single particle dominates all the other particles that can not provide further information. The re-sampling multiplies the particles with probability masses corresponding to the weights and looses the particles that have small weights. The re-sampled set of particles contains evenly distributed weights, which can lead the filter to a worse state than before due to random process involved in the re-sampling. A threshold value can be set to decide when to re-sample the particles by calculating the effective sample size (ESS) with the following equation:

$$ESS(w_t) \triangleq \left(\sum_{i=1}^{N} {w_t^{(i)}}^2 \right)^{-1} \tag{1.18}$$

which is equal to maximum of N when all the weights are evenly distributed, and equal to minimum of 1 when a single particle has all the probability mass. The ESS is useful for independent particles and, thus, it is computed before the re-sampling. At each

time step, the ESS is calculated and compared to a threshold size and the particles are re-sampled if it is below the threshold.

1.2. Experimental Setup

This section demonstrates the experimental setup of the proposed study, introduces the magnetically actuated soft capsule endoscope (MASCE) and explains how the training and testing datasets were recorded.

1.2.1. Magnetically Actuated Soft Capsule Endoscopes (MASCE)

The capsule prototype used in the experiments is a magnetically actuated soft capsule endoscope (MASCE) designed for disease detection, drug delivery and biopsy operations in the upper GI-tract. The prototype is composed of a RGB camera, a permanent magnet, an empty space for drug chamber and a biopsy tool (see Figure 2.7 and 1.7 for visual reference). The magnet exerts magnetic force and torque to the robot in response to a controlled external magnetic field [1]. The magnetic torque and forces are used to actuate the capsule robot and to release drug. Magnetic fields from the electromagnets generate the magnetic force and torque on the magnet inside MASCE so that the robot moves inside the workspace. Sixty-four three-axis magnetic sensors are placed on the top, and nine electromagnets are placed in the bottom [1].

1.2.2. Magnetic Localization System

Our 5-DoF magnetic localization system is designed for the position and orientation estimation of untethered meso-scale magnetic robots [22]. The system, which is described in Figure 1.6, uses external magnetic sensor system and electromagnets for the localization of the magnetic capsule robot. A 2D-Hall-effect sensor array measures magnetic field from the magnetic capsule robot outside the workspace. Additionally, computer-controlled magnetic coil array consisting of nine electromagnets generates actuator's magnetic field. The core idea of our localization technique is separation of capsule's magnetic field from actuator's magnetic field. For that purpose, actuator's

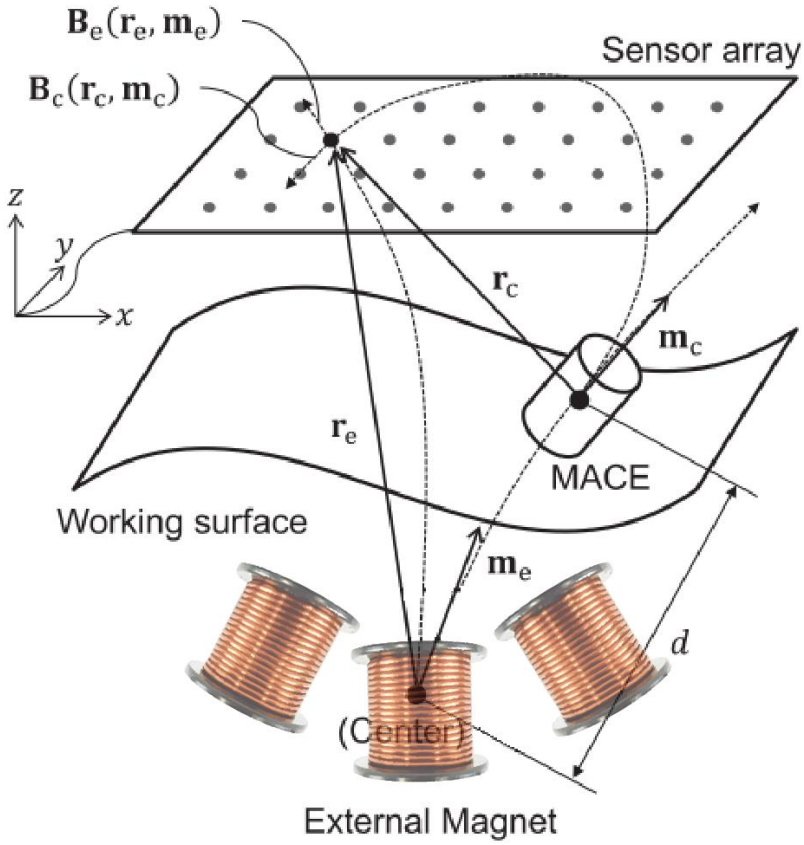


Figure 1.6. Schematic drawing of the magnetic localization technique. A capsule robot is manipulated by an external magnetic coil array consisting of nine electromagnets [1].

magnetic field is subtracted from the magnetic field data which is acquired by Hall-effect sensor array. As a further step, second-order directional differentiation is applied to reduce the localization error [22].

1.2.3. Monocular Visual Odometry

For every input RGB image, we create its depth image using the source code of the perspective shape-from-shading under realistic lighting conditions project [44]. Once perspective shape from shading based depth map for the input image is obtained, the framework uses both RGB and depth map information to jointly estimate camera pose. An energy minimization based pose estimation technique is applied containing both sparse optical flow (OF) based correspondence establishment, and volumetric and

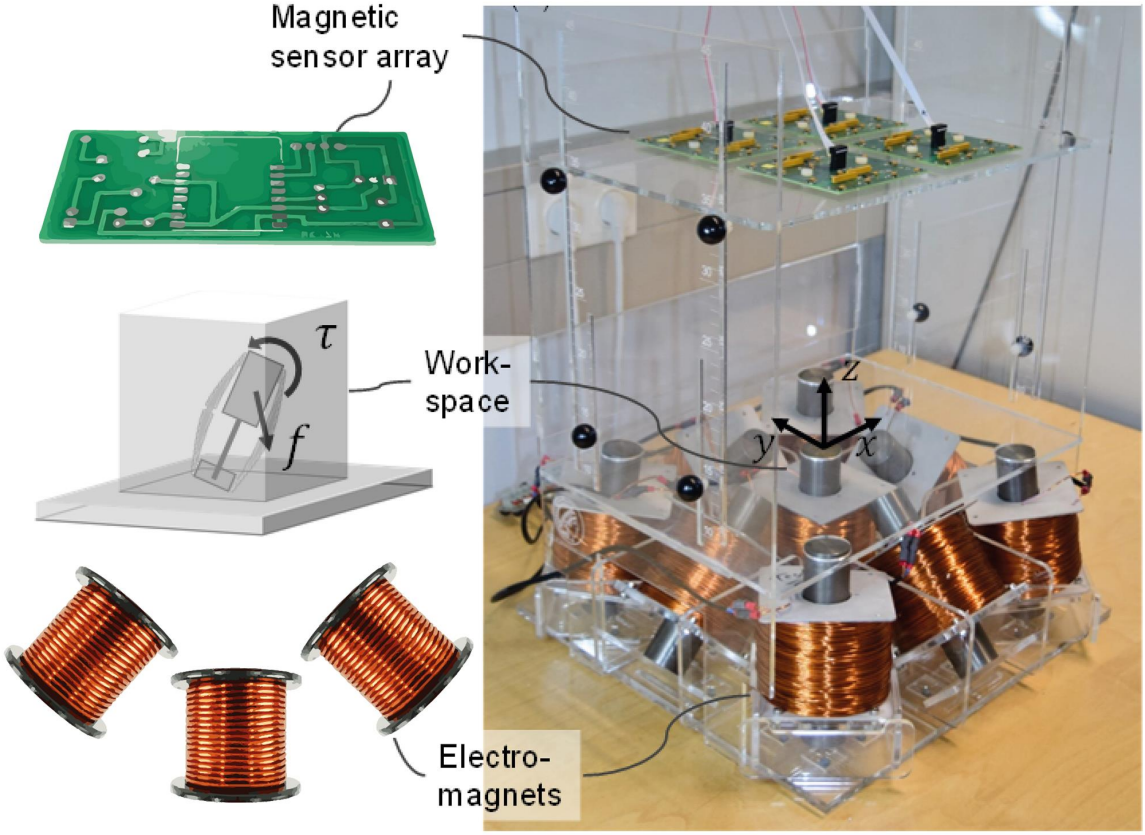


Figure 1.7. Actuation system of the MASCE [1].

photometric dense alignment establishment [45–47]. Inspired from the pose estimation strategies proposed by [45–47], for a parameter vector

$$X = (R_o, t_o, ..., R_{|S|}, t_{|S|})^T \tag{1.19}$$

for $|S|$ frames, the alignment problem is defined as a variational non-linear least squares minimization problem with the following objective, consisting of the OF based pixel correspondences and dense jointly photometric-geometric constraints [45–47]. Outliers after OF estimation are eliminated using motion bounds criteria, which removes pixels with a very large displacement and too different motion vector than neighbouring pixels. The energy minimization equation is as follows:

$$E_{\text{align}}(X) = \omega_{\text{sparse}} E_{\text{sparse}}(X) + \omega_{\text{dense}} E_{\text{dense}}(X) \tag{1.20}$$

Table 1.1. Endoscopic camera specifications used for the experiments.

(a) Awaiba Naneye Endoscopic Camera.

	Specifications
Resolution	250 x 250 pixel
Footprint	2.2 x 1.0 x 1.7 mm
Pixel size	3 x 3 μm^2
Frame rate	44 fps

(b) Misumi-V3506-2ES camera.

	Specifications
Resolution	400 x 400 pixel
Diameter	8.2mm
Pixel size	5.55 x 5.55 μm^2
Frame rate	30 fps

(c) Misumi-V5506-2ES camera.

	Specifications
Resolution	640 x 480 pixel
Diameter	8.6 mm
Pixel size	6.0 x 6.0 μm^2
Frame rate	30 fps

(d) Potensic Mini Camera.

	Specifications
Resolution	1280 x 720 pixel
Diameter	8.8 mm
Pixel size	10.0 x 10.0 μm^2
Frame rate	30 fps

where, ω_{sparse} and ω_{dense} are weights assigned to sparse and dense matching terms and E_{sparse} (X) and $E_{\text{dense}}(X)$ are the sparse and dense matching terms, respectively, such that:

$$E_{\text{sparse}}(X) = \sum_{\text{i=1}}^{|S|} \sum_{\text{j=1}}^{|S|} \sum_{\text{(k,1)} \in \text{C(i,j)}} ||\tau_i P_{\text{i,k}} - \tau_j P_{\text{j,k}}||^2 \tag{1.21}$$

Here, $P_{\text{i, k}}$ is the k^{th} detected feature point in the i-th frame. $C(i, j)$ is the set of all pairwise correspondences between the i-th and the j-th frame. The Euclidean distance over all the detected feature matches is minimized once the best rigid transformation τ_i is found. Dense pose estimation is described as follows [45–47]:

$$E_{\text{dense}}(\tau) = \omega_{\text{photo}} E_{\text{photo}}(\tau) + \omega_{\text{geo}} E_{\text{geo}}(\tau) \tag{1.22}$$

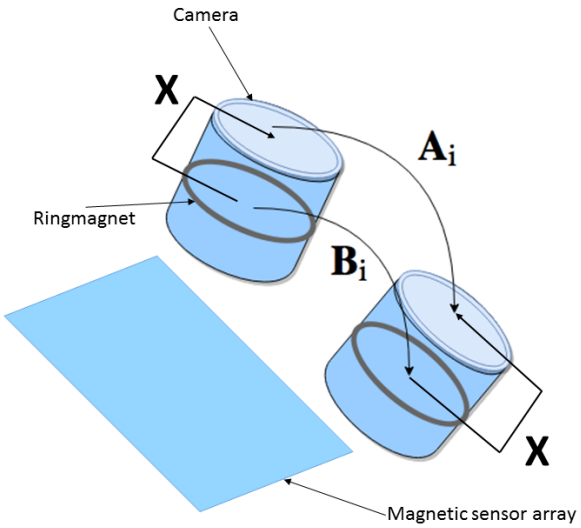


Figure 1.8. The transformations between different frames at pose i and pose $i+1$ [2].

where,

$$E_{\text{photo}}(X) = \sum_{\text{(i,j)} \in \text{E}} \sum_{\text{k=0}}^{|I_i|} ||I_i(\omega(d_{\text{i,k}})) - I_j(\omega(\tau_j^{-1}\tau_i d_{\text{i,k}}))||_2^2 \tag{1.23}$$

and,

$$E_{\text{geo}}(X) = \sum_{\text{(i,j)} \in \text{E}} \sum_{\text{k=0}}^{|D_i|} [n_{\text{i,k}}^T (d_{\text{i,k}} - \tau_i^{-1}\tau_j \omega^{-1}(D_j(\omega(\tau_j^{-1}\tau_i d_{\text{i,k}}))))]^2 \tag{1.24}$$

with τ_i being rigid camera transformation, $P_{\text{i,k}}$ the k^{th} detected inlier point in i^{th} frame, and $C(i,j)$ being the set of pairwise correspondences between the i^{th} and j^{th} frame. In Equation 1.20, ω_{dense} is linearly increased; this allows the sparse term to first find a good global structure, which is then refined with the dense term (coarse-to-fine alignment [45]). Using Gauss-Newton optimization, we find the best pose parameters X which minimizes the proposed highly non-linear least squares objective. for further details, the reader is referred to [2].

1.2.4. Hand-eye Calibration

To relate measurements made by the magnetic sensor array and capsule camera, first the transformation between the coordinate systems of both sensors has to be estimated. For that purpose, we make use of dual quaternion based hand-eye calibration method proposed by [48], which relies on the invariance of the angle and the pitch provided by the dual quaternion parametrization. We denote by $\mathbf{X}$ the transformation from magnetic sensor to capsule camera, by $\mathbf{A}_i$ the transformation matrix between camera frame i and frame $i+1$, and by $\mathbf{B}_i$ the transformation matrix from ring magnet pose i to ring magnet pose $i+1$. Figure 1.8 illustrates the schema of our sensor-to-sensor transformation approach. For solving the transformation problem, we collect 20 different capsule robot pose values, which gives us 20 different instances of the equation $\mathbf{AX} = \mathbf{XB}$ with the unknown transformation matrix $\mathbf{X}$. Magnetic sensor delivers 50 Hz data, whereas camera has 30 fps data frequency. For a tight synchronization of camera and magnetic sensor outputs, we use global world clock of the operation PC, where a very tiny difference of milliseconds still remains, which is solved by interpolation of values in-between. Since the magnetic localization system is 5-DoF and visual odometry delivers 6-DoF localization, we assign for the 6th degree of the magnetic localization the angle 0° as it is an independent rotational parameter since it has no influence on the other 5-DoF. Using SVD (Singular Value Decomposition), we solve the equations and determine the magnetic-to-camera transformation matrix $\mathbf{X}$.

2. MULTI-SENSOR FUSION WITH SWITCHING STATE-SPACE MODEL

The statistical Bayesian filtering method is employed to compute the posterior probability density function (pdf) of observations or sequentially obtained state vectors $\mathbf{x}_t \in \mathcal{X}$ of sensor measurements. The model can be described as:

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{v}_t) \tag{2.1}$$

where f is a non-linear state transition function and $\mathbf{v_t}$ is a white noise. t is the index of a time sequence, $t \in \{1, 2, 3, ...\}$.

6-DoF pose state estimation with a high precision is a complex problem, which often requires multi-sensor input or sequential observations. In our capsule, we have two sensor systems, one being a 5-DoF magnetic sensor array and the other one being an endoscopic monocular RGB camera (these subsystems are described later). Those observations can be represented as $\mathbf{z}_{k,t}(k = 1, ..., n)$, where n is the number of sensors with the probability of each observation $p(\mathbf{z}_{k,t}|\mathbf{x}_t)$.

2.1. The Sequential Bayesian Model and Problem Statement

We estimate the 6-DoF pose states which rely on latent (hidden) variables by using the Bayesian filtering approach. The statistical conditional relations between all of the variables is shown in the probabilistic graphical model (see Figure 2.1). The hidden variables of sensor states are denoted as $s_{k,t}$, which we call switch variable, where $s_{k,t} \in \{0, ..., d_k\}$ for $k = 1, ..., n$. d_k is the number of possible observation models, e.g., failure and nominal sensor states. The observation model for $\mathbf{z}_{k,t}$ can be described as:

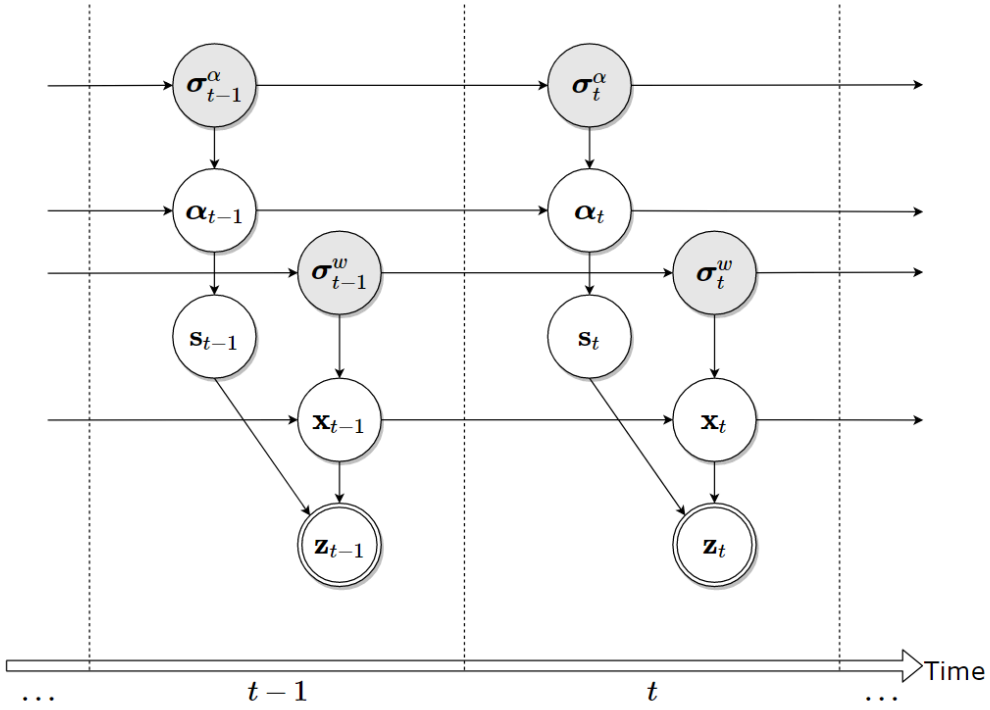


Figure 2.1. The overall switching state-space model. Observable variables and hyper-parameters are denoted by double circles and gray circles, respectively.

$$\mathbf{z}_{k,t} = h_{k,s_{k,t},t}(\mathbf{x}_t) + \mathbf{w}_{k,s_{k,t},t} \tag{2.2}$$

where $h_{k,s_{k,t},t}(\mathbf{x}_t)$ is the non-linear observation function and $\mathbf{w}_{k,s_{k,t},t}$ is the observation noise. The latent variable of the switch parameter $s_{k,t}$ is defined to be 0 if the sensor is in a failure state, which means that observation $\mathbf{z}_{k,t}$ is independent of $\mathbf{x}_t$, and 1 if the sensor k is in its nominal state of work. The prior probability for the switch parameter $s_{k,t}$ being in a given state j, is denoted as $\alpha_{k,j,t}$ and it is the probability for each sensor to be in a given state:

$$Pr(s_{k,t} = j) = \alpha_{k,j,t}, \quad 0 \leq j \leq d_k \tag{2.3}$$

where $\alpha_{k,j,t} \geq 0$ and $\sum_{j=0}^{d_k} \alpha_{k,j,t} = 1$ with a Markov evolution model. The objective posterior pdf $p(\mathbf{x}_{0:t}, \mathbf{s}_{1:t}, \alpha_{0:t}|\mathbf{z}_{1:t})$ and the marginal posterior probability $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, in general, cannot be determined in a closed form due to its complex shape. However, sequential Monte Carlo methods (*particle filters*) provide a numerical approximation of the posterior pdf with a set of samples (*particles*) weighted by the kinematics and observation models.

2.2. Proposal Distributions

In this section, we formulate the optimal proposal distributions in terms of minimizing the variance of the weights and effective approximations, in cases where sampling from the optimal distributions is not feasible. The particles are extended from time $t-1$ to time t according to the proposal distribution denoted by $q(\cdot)$.

- $q(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(i)}, \sigma_t^{w(i)}, \hat{\mathbf{s}}_t^{(i)}, \mathbf{z}_t)$ is approximated by an unscented Kalman filter (UKF) step:

$$\hat{\mathbf{x}}_{t|t}^{(i)} = \hat{\mathbf{x}}_{t|t-1}^{(i)} + \sum_{k=1}^{n} \hat{s}_{k,t}^{(i)} K_{k,t}^{(i)} \hat{\nu}_{k,t}^{(i)}$$

  where $\hat{\mathbf{x}}_{t|t-1}^{(i)} = f(\mathbf{x}_{t-1}^{(i)})$, n is the number of sensors, $\hat{\nu}_{k,t}^{(i)}$ is the residual, and $K_{k,t}^{(i)}$ is the Kalman gain sequentially obtained by UKF. Finally,

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(i)}, \sigma_t^{w(i)}, \hat{\mathbf{s}}_t^{(i)}, \mathbf{z}_t) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}^{(i)}, P_{t|t}^{(i)})$$

  where the error covariance matrix, $P_{t|t}^{i}$ is obtained by the UKF step with the process noise of $\sigma_t^{w(i)}$.
- In switching state-space models, the switch parameters with self-adaptive prior are more efficient than a fixed prior approach [49, 50]. The optimal proposal distribution for switch variable that represents the state of a sensor is given by

$$Pr(\mathbf{s}_{k,t}|\mathbf{x}_{t-1}^{(i)}, \alpha_{k,t-1}^{(i)}, \mathbf{z}_{k,t}) = \frac{\alpha_{k,s_{k,t},t-1}^{(i)} p(\mathbf{z}_{k,t}|s_{k,t}, \mathbf{x}_{t-1}^{(i)})}{\sum_{j=0}^{d_k} \alpha_{k,s_{k,t},t-1}^{(i)} p(\mathbf{z}_{k,t}|j, \mathbf{x}_{t-1}^{(i)})} \tag{2.4}$$

which is approximated by applying UKF to pdfs $p(\mathbf{z}_{k,t}|j, \mathbf{x}_{t-1}^{(i)})$ for $j = 0, ..., d_k$

$$p(\mathbf{z}_{k,t}|j, \mathbf{x}_{t-1}^{(i)}) \simeq \mathcal{N}(h_{k,j,t}(\hat{x}_{t|t-1}^{(i)}), S_{k,j,t}^{(i)}) \tag{2.5}$$

where $\hat{x}_{t|t-1}^{(i)} = f(\mathbf{x}_{t-1}^{(i)})$ is the state prediction and $S_{k,j,t}^{(i)}$ is the approximated innovation covariance matrix approximated by UKF. Hence, the proposal distribution for the switch parameter $\mathbf{s}_{k,t}$ is given by

$$q\big(\mathbf{s}_{k,t}|\mathbf{x}_{t-1}^{(i)}, \alpha_{k,t-1}^{(i)}, \mathbf{z}_{k,t}\big) \propto \alpha_{k,s_k,t-1}^{(i)} \mathcal{N}(h_{k,s_k,t-1}(\hat{x}_{t|t-1}^{(i)}), S_{k,s_k,t}^{(i)}) \tag{2.6}$$

- The optimal proposal distribution for the hyperparameter $\sigma_{k,t-1}^{\alpha}$ is calculated in closed form as

$$\begin{aligned} &q\bigg(\log(\sigma_{k,t}^{\alpha})|\alpha_{k,t}^{(i)}, \alpha_{k,t-1}^{(i)}, \sigma_{k,t-1}^{\alpha(i)} \bigg) \\ &= \frac{D\bigg(\alpha_{k,t}^{(i)}; \sigma_{k,t}^{\alpha} \alpha_{k,t-1}^{(i)} \bigg)}{D\bigg(\alpha_{k,t}^{(i)}; \sigma_{k,t-1}^{\alpha} \alpha_{k,t-1}^{(i)} \bigg)} \times \mathcal{N}\bigg(\log(\sigma_{k,t}^{\alpha}); \log(\sigma_{k,t-1}^{\alpha(i)}), \lambda^{\alpha} \bigg). \end{aligned} \tag{2.7}$$

We generate samples from the distribution with Adaptive Rejection Sampling (ARS) method since direct sampling is not feasible [51]. Using ARS, the need for locating the supremum diminishes because the distribution is log-concave. Another advantage of ARS is that it uses recently acquired information to update the envelope and squeezing functions, which reduces the need to evaluate the distribution after each rejection step. Figure 2.2 shows an ARS sampling result indicating the effectiveness of the applied sampling method for the proposal dis-

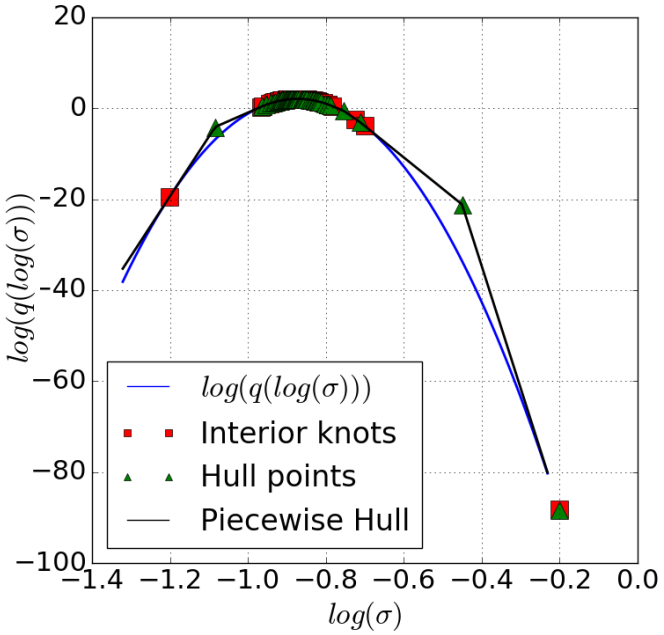


Figure 2.2. Example ARS sampling result for $\log(\sigma_{k,t})$. The piecewise hull and the generated samples are shown.

tribution. It can be seen in Figure 2.2 that a tight piecewise hull has converged to the target distribution after rejection steps and interior knots are regenerated in the vicinity of the expected values.

- Considering that the Dirichlet distribution is conjugate to the multinomial distribution, the optimal proposal distribution for the confidence parameter $\alpha_{k,t}$ can be reformulated in a closed form as a Dirichlet distribution with a decreasing variance parameter for failure sensor states.

2.3. The Particle Filter Algorithm

The synchronized sensor measurements delivered by the visual odometry system and the magnetic localization system are used to sequentially update the proposed multi-sensor fusion model. The corresponding particle filter method with the initialization, sequential update and the re-sampling parts is described in Figure 2.3, Figure 2.4 and Figure 2.5, respectively.

Initialization

for $i = 1$ to N **do**

Sample: $x_0^{(i)} \sim p_0(x_0)$

Sample: *sample* $\sigma_0^{(i)} \sim p_0(\sigma_0)$

Sample: *sample* $\alpha_0^{(i)} \sim p_0(\alpha_0|\sigma_0^{(i)})$

Initial weights: $w_0^{(i)} \leftarrow (1/N)$

end for

Figure 2.3. Initialization of the particle filter algorithm used in the multi-sensor fusion model.

2.4. RNN-based Kinematics Model

Existing sensor fusion methods based on traditional particle filter and Kalman filter approaches have their limitations when applied to nonlinear dynamic systems. The Kalman filter and extended Kalman filter assume that the underlying dynamic process is well-modeled by linear equations or that these equations can be linearised without a major loss of fidelity. On the other hand, particle filters accommodate a wide variety of dynamic models, allowing for highly complex dynamics in the state variables. In the last years, deep learning (DL) techniques have been dominating many computer vision and sequential learning related tasks with promising results, e.g., object detection, object recognition, classification problems, time series estimations, natural language processing etc. Contrary to these high-level tasks, multi-sensory data fusion is mainly working on motion dynamics and relations across sequence of pose observations obtained from sensors, which can be formulated as a sequential learning problem. Unlike traditional feed-forward artificial neural networks, RNNs are very suitable for modelling the dependencies across time sequences and for creating a temporal motion model since it has a memory of hidden states over time and has directed cycles among hidden units, enabling the current hidden state to be a function of arbitrary sequences of inputs (see Figure 2.6). Thus, using RNN, the pose estimation of the current time step benefits from information encapsulated in previous time steps [52],

Sequential updates:

for $t = 1$ to T **do**

  for $i = 1$ to N **do**

    Sample the sensor state variable $\tilde{s}_t^{(i)} \sim q(\mathbf{s}_t|\mathbf{x}_{t-1}^{(i)}, \alpha_{t-1}^{(i)}, \mathbf{z}_t)$

    Sample the state vector $\tilde{x}_t^{(i)} \sim q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}, \tilde{s}_{t-1}^{(i)}, \mathbf{z}_t)$

    Sample the probabilities $\tilde{\alpha}_t^{(i)} \sim q(\alpha_t|\alpha_{t-1}^{(i)}, \tilde{s}_{t-1}^{(i)}, \sigma_{t-1}^{(i)})$

    Sample the hyperparameter vector $\tilde{\sigma}_t^{(i)} \sim q(\sigma_t|\sigma_{t-1}^{(i)}, \tilde{\alpha}_t^{(i)}, \alpha_{t-1}^{(i)})$

  end for

  for $i = 1$ to N **do**

    Weight update: Sequentially update the weights of the particles with the following equation:

$$\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(\mathbf{z}_t|\tilde{\mathbf{x}}_t^{(i)}, \tilde{s}_t^{(i)}) p(\tilde{x}_t^{(i)}|\tilde{x}_{t-1}^{(i)})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}, \tilde{s}_{t-1}^{(i)}, \mathbf{z}_t) q(\mathbf{s}_t|\mathbf{x}_{t-1}^{(i)}, \alpha_{t-1}^{(i)}, \mathbf{z}_t)} \times \frac{p(\tilde{\mathbf{s}}_t^{(i)}|\tilde{\alpha}_t^{(i)}) p(\tilde{\alpha}_t^{(i)})|\alpha_{t-1}^{(i)}, \tilde{\sigma}_t^{(i)}|\tilde{\sigma}_{t-1}^{(i)}}{q(\tilde{\alpha}_t^{(i)}|\tilde{\alpha}_{t-1}^{(i)}) q(\sigma_t|\sigma_{t-1}^{(i)}, \tilde{\alpha}_t^{(i)}, \alpha_{t-1}^{(i)})} \tag{2.8}$$

    Weight normalization: $\sum_{i=1}^{N} \tilde{w}_t^{(i)} = 1$

  end for

end for

Figure 2.4. Sequential updates of the particle filter algorithm used in the multi-sensor fusion model.

Resampling

Compute $ESS(\mathbf{w}_t)$ and decide a threshold η, e.g., $\eta = 0.8 \times N$

if $ESS(\mathbf{w}_t) \leq \eta$ **then**

Re-sample the particles.

Reset the weights of the resulting particles $\mathbf{x}_t^{(i)}$ with $w_t^{(i)} = 1/N$

else

Rename the particles: $\mathbf{x}_t^{(i)} \leftarrow \tilde{\mathbf{x}}_t^{(i)}$

end if

Figure 2.5. Re-sampling method of the particle filter algorithm used in the multi-sensor fusion model.

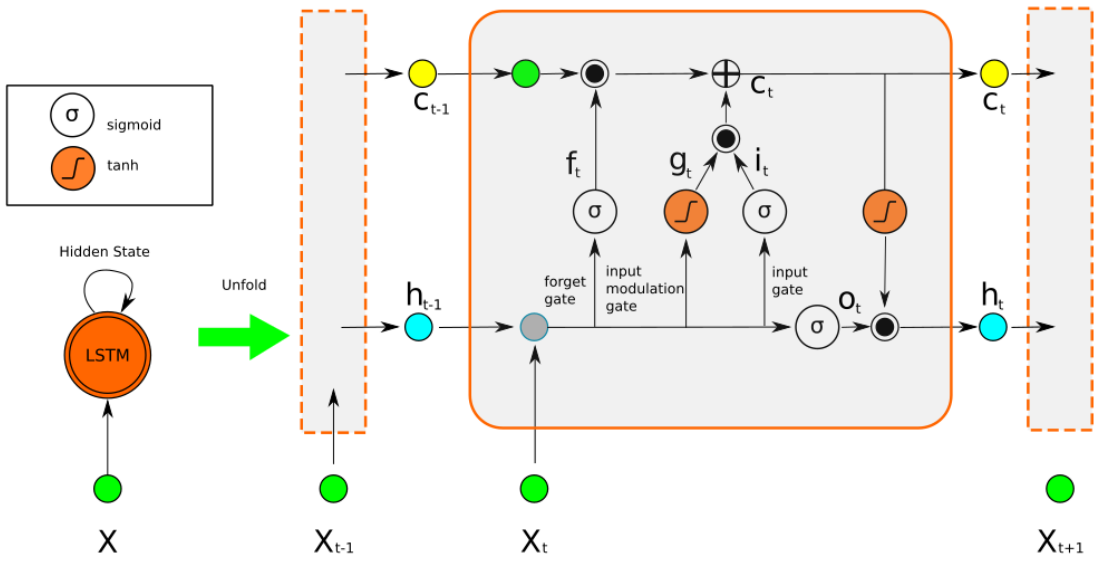


Figure 2.6. Data flow through the hidden units of the LSTM [3].

which is suitable to formulate the state transition function f in Equation 2.1.

The training data is divided into input sequences of length 50 and the slices are passed into the RNN modules with an expectation that RNN predicts the next 6-DoF pose value. Long Short-Term Memory (LSTM) is a suitable implementation of RNN to exploit longer trajectories since it avoids the vanishing gradient problem of RNN resulting in a higher capacity of learning long-term relations among the sequences by introducing memory gates such as input, forget and output gates, and hidden units of several blocks. The input gate controls the amount of new information flowing into the current state, the forget gate adjusts the amount of existing information that remains in the memory, and the output gate decides which part of the information triggers the activations. The folded LSTM and its unfolded version over time are shown in Figure 2.6 along with the internal structure of a LSTM memory cell. Given the input vector x_k at time k, the output vector h_{k-1} and the cell state vector c_{k-1} of the previous LSTM unit, the LSTM updates at time step k according to the following equations:

$$
\begin{aligned}
f_k &= \sigma(W_f \cdot [x_k, h_{k-1}] + b_f) \\
i_k &= \sigma(W_i \cdot [x_k, h_{k-1}] + b_i) \\
g_k &= tanh(W_g \cdot [x_k, h_{k-1}] + b_g) \\
c_k &= f_k \odot c_{k-1} + i_k \odot g_k \\
o_k &= \sigma(W_o \cdot [x_k, h_{k-1}] + b_o) \\
h_k &= o_k \odot tanh(c_k)
\end{aligned}
$$

where σ is sigmoid non-linearity, tanh is hyperbolic tangent non-linearity, W terms denote corresponding weight matrices, b terms denote bias vectors, $\odot$ is the Hadamard product, i_k, f_k, g_k, c_k and o_k are input gate, forget gate, input modulation gate, the cell state and output gate at time k, respectively [3].

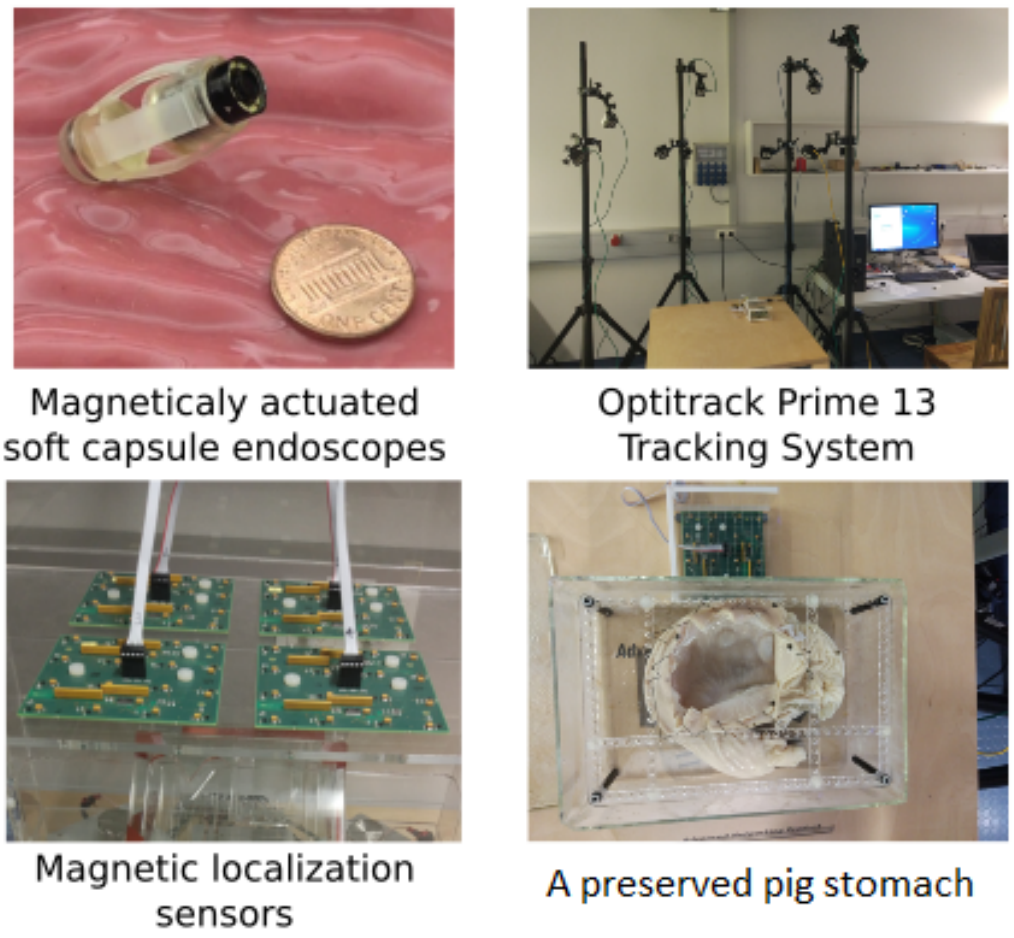


Figure 2.7. Experimental setup [4].

The back-propagation algorithm is used to calculate the gradients of RNN weights, which are passed to the Adam optimization method to compute adaptive learning rates for each parameter employing the first-order gradient-based optimization of the stochastic objective function [53]. In addition to saving exponentially decaying average of past squared gradients, v_t, Adam optimization keeps exponentially decaying average of past gradients, m_t that is similar to momentum. The update equations are given as

$$(m_t)_i = \beta_1(m_{t-1})_i + (1-\beta_1)(\nabla L(W_t))_i \tag{2.9}$$

$$(v_t)_i = \beta_2(v_{t-1})_i + (1-\beta_2)(\nabla L(W_t))_i^2 \tag{2.10}$$

$$(W_{t+1})_i = (W_t)_i - \alpha \frac{\sqrt{1-(\beta_2)_i^t}}{1-(\beta_1)_i^t} \frac{(m_t)_i}{\sqrt{(v_t)_i + \varepsilon}} \tag{2.11}$$

where we used default values proposed by [53] for the parameters β_1, β_1 and ε: $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$.

2.5. Dataset

We created a training dataset, which was recorded on five different real pig stomachs (see Figure 2.7). To ensure that our algorithm is not tuned to a specific camera model, four different commercial endoscopic cameras were employed. For each pig stomach-camera combination, $2,000$ frames were acquired which makes for four cameras and five pig stomachs $40,000$ frames, in total. Sample real pig stomach frames are shown in Figure 2.8 for the visual reference. During video recording, Optitrack motion tracking system consisting of eight Prime-13 cameras and a tracking software was utilized to obtain 6-DoF localization ground truth data in a sub-millimeter precision (see Figure 2.7), which was used as a gold standard for the evaluations of the pose estimation accuracy. We divided our dataset into two groups. First group consisting of $30,000$ frames was used for RNN training purposes, whereas the last $10,000$ frames

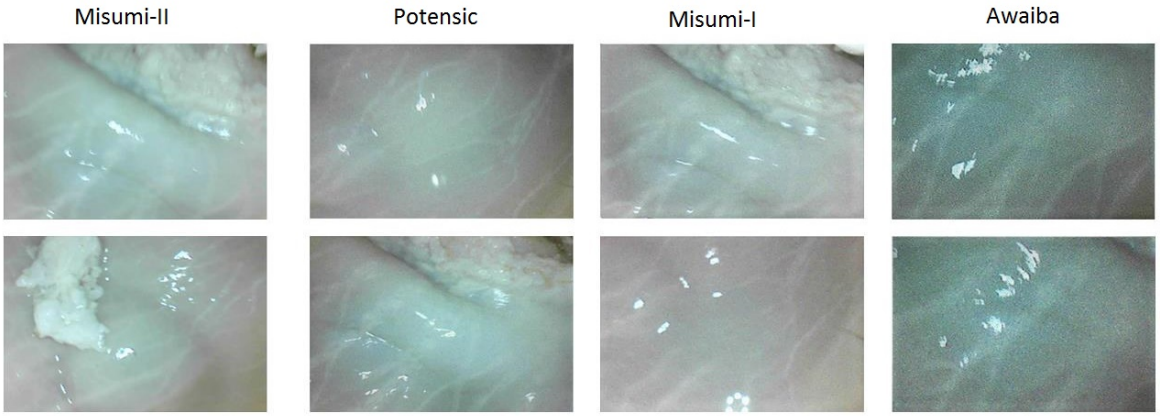


Figure 2.8. Sample frames from the dataset used in the experiments [4].

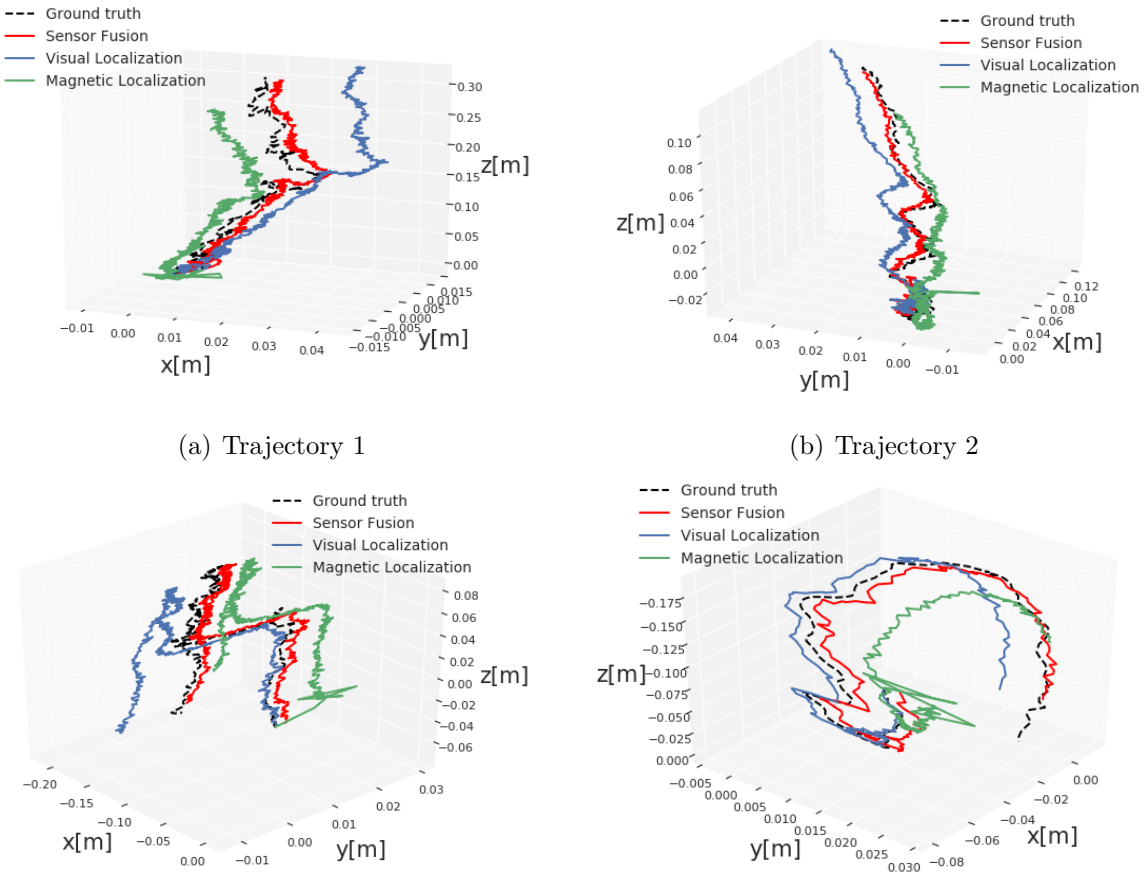
were used for testing, which were not used for the training section.

3. EXPERIMENTS AND RESULTS

LSTM module was trained using Keras library with GPU programming and Theano back-end. Using back-propagation-through-time method, the weights of hidden units were trained for up to 200 epochs with an initial learning rate of 0.001. Overfitting was prevented using dropout and early stopping techniques. Dropout regularization technique introduced by [54] is an extremely effective and simple method to avoid overfitting. It samples a part of the whole network and updates its parameters based on the input data. Early stopping is another widely used technique to prevent overfitting of a complex neural network architecture which was optimized by a gradient-based method.

The performance of the proposed multi-sensor fusion approach was analysed by examining posterior probabilities of the switch parameters $\mathbf{s}_{k,t}$ (see Figure 3.2), the minimum mean square error (MMSE) estimates of $\alpha_{k,t}$ (see Figure 3.3) and evolution of the hyper-parameter $\sigma^{\alpha}_{k,t}$ (see Figure 3.4). Moreover, for various trajectories with different complexity levels of motions, including uncomplicated paths with slow incremental translations and rotations, comprehensive scans with many local loop closures and complex paths with sharp rotational and translational movements, we analysed both the localization accuracy and the fault detection performance of our multi-sensor fusion approach (see Figure 3.1 and 3.5). Additionally, we compared the localization accuracy of the multi-sensor fusion approach with the visual localization and magnetic localization (see Figure 3.5) using RMSE.

The results in Figure 3.2 and Figure 3.3 indicate that the sensor states are accurately estimated (visual localization fails between 14-36 seconds and magnetic sensor fails between 57-76 seconds. Both failures are detected successfully.), and the MMSE is kept low, thanks to the switching option ability from one observation model to another in case of a sensor failure. In our model, we do not make a Markovian assumption for the switch variable $\mathbf{s}_{k,t}$ but we do for its prior $\alpha_{k,t}$, resulting in a priori dependent on the past trajectory sections, which is more likely for the incremental endoscopic



(a) Trajectory 1

(b) Trajectory 2

(c) Trajectory 3

(d) Trajectory 4

Figure 3.1. Sample trajectories comparing the multi-sensor fusion result with ground truth and sensor data.

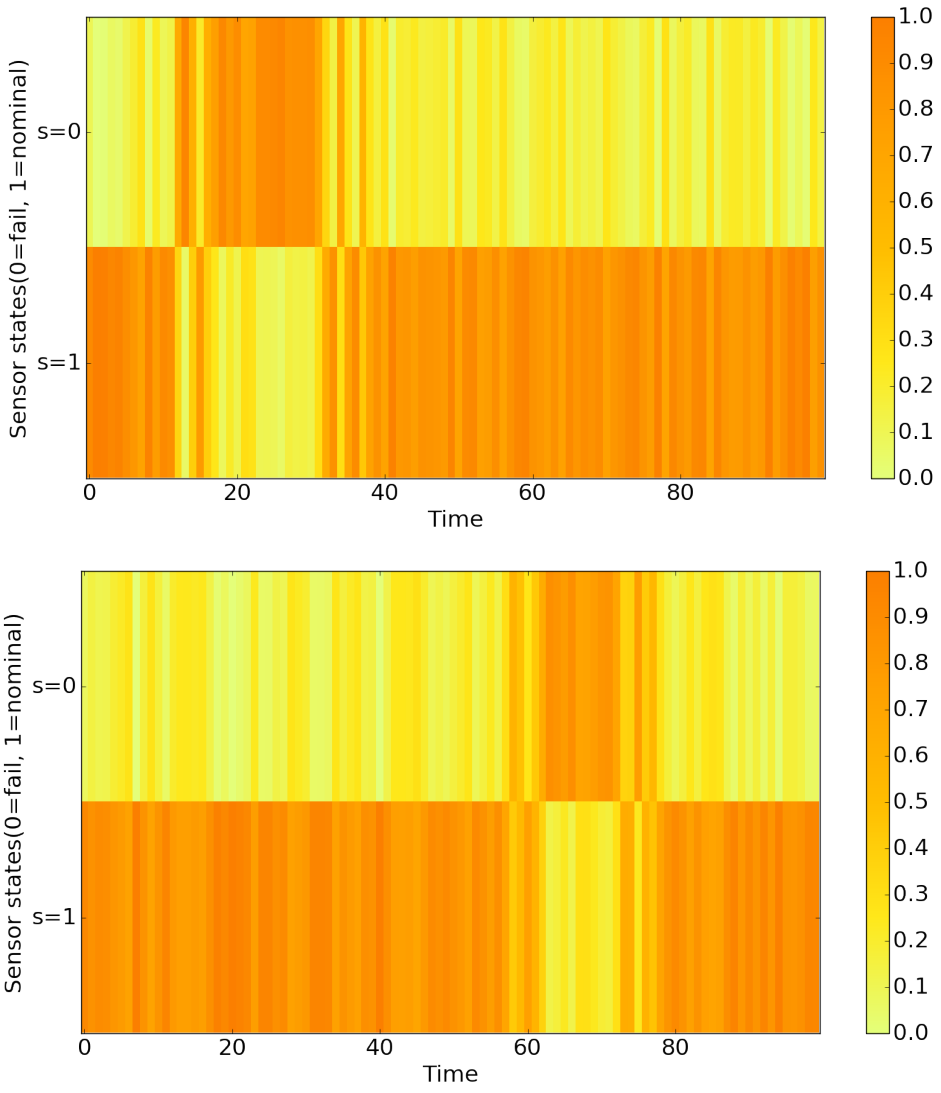


Figure 3.2. Posterior probability of $\mathbf{s}_{k,t}$ parameter for endoscopic RGB camera (top) and for magnetic localization system (bottom). The switch parameter, $\mathbf{s}_{k,t}$, reflects the failure times accurately.

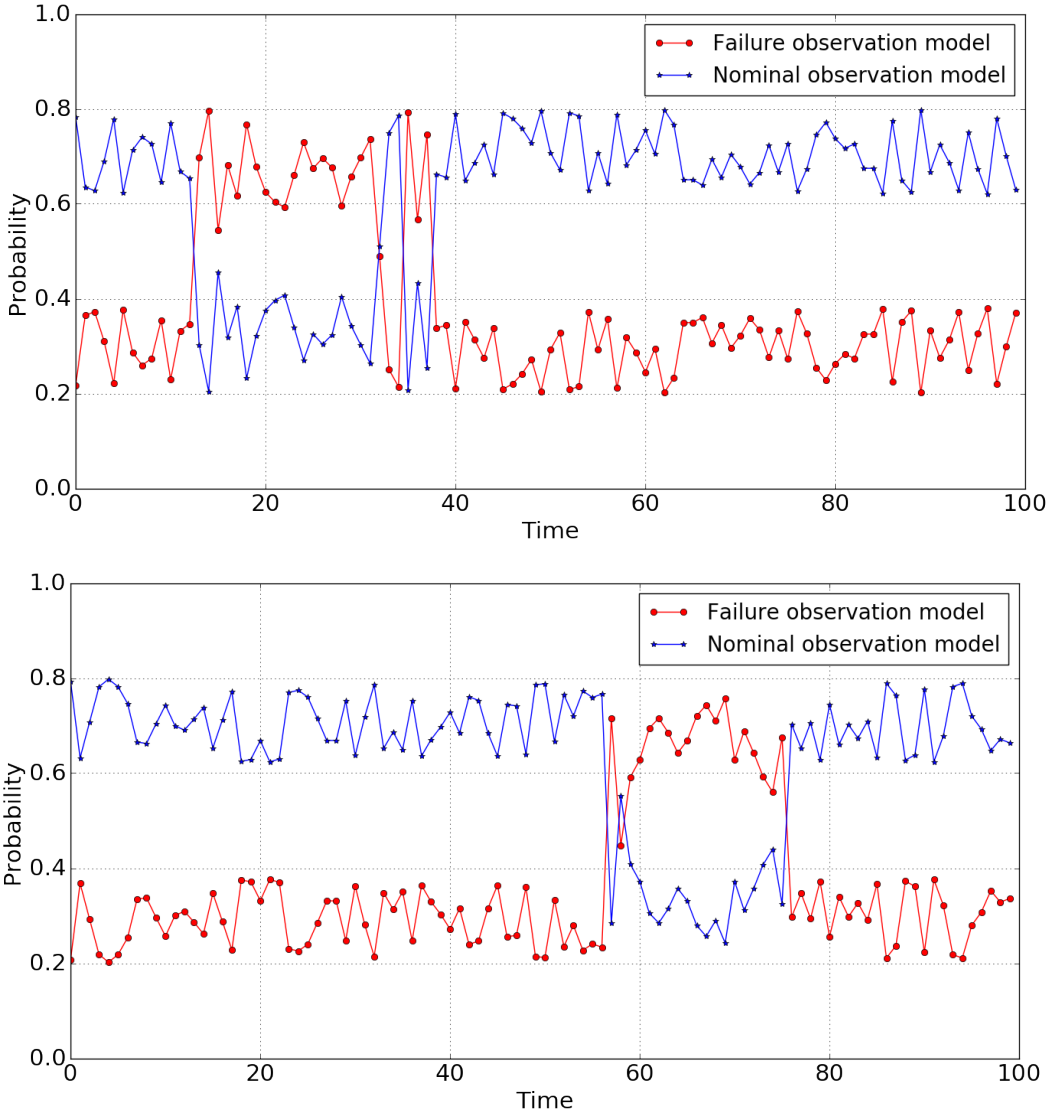


Figure 3.3. The minimum mean square error (MMSE) of $\alpha_{k,t}$ for endoscopic RGB camera (top) and for magnetic localization system (bottom).

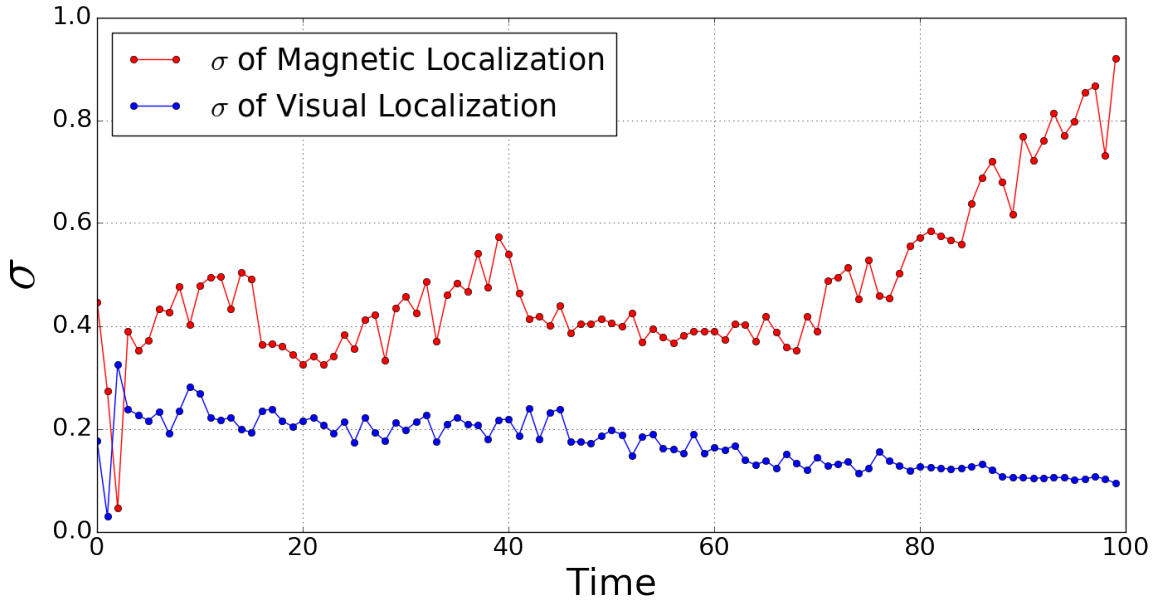


Figure 3.4. Evolution of the $\sigma^{\alpha}_{k,t}$ parameter for the sensors. $\sigma^{\alpha}_{k,t}$ does not tend to increase during sensor failure periods.

capsule robot motions. Our model thus introduces a memory over the past sensor states rather than simply considering the last state. The length of the memory is tuned by the hyper-parameters $\sigma^{\alpha}_{k,t}$, leading to a long memory for large values and vice-versa. This is of particular interest when considering sensor failures. Our system detects automatically failure states. Thus, the confidence in the RGB sensor decrease when visual localization fails recently due to occlusions, fast-frame-to frame changes etc. On the other hand, the confidence in magnetic sensor decreases if the magnetic localization fails due to noise interferences from environment or if the ringmagnet has a big distance to the magnetic sensor array.

The results depicted in Figure 3.1 indicate that the proposed model clearly outperforms magnetic and visual localization approaches in terms of translational and rotational pose estimation accuracy. The multi-sensor fusion approach is able to stay close to the ground truth pose values for even sharp crispy motions despite sensor failures. Even for very fast and challenging paths that can be seen in Figure 3.1(c) and 3.1(d), the deviations of sensor fusion approach from the ground-truth still remain in an acceptable range for medical operations. We presume that the effective use of switching observations and particle filtering with non-linear motion estimation using

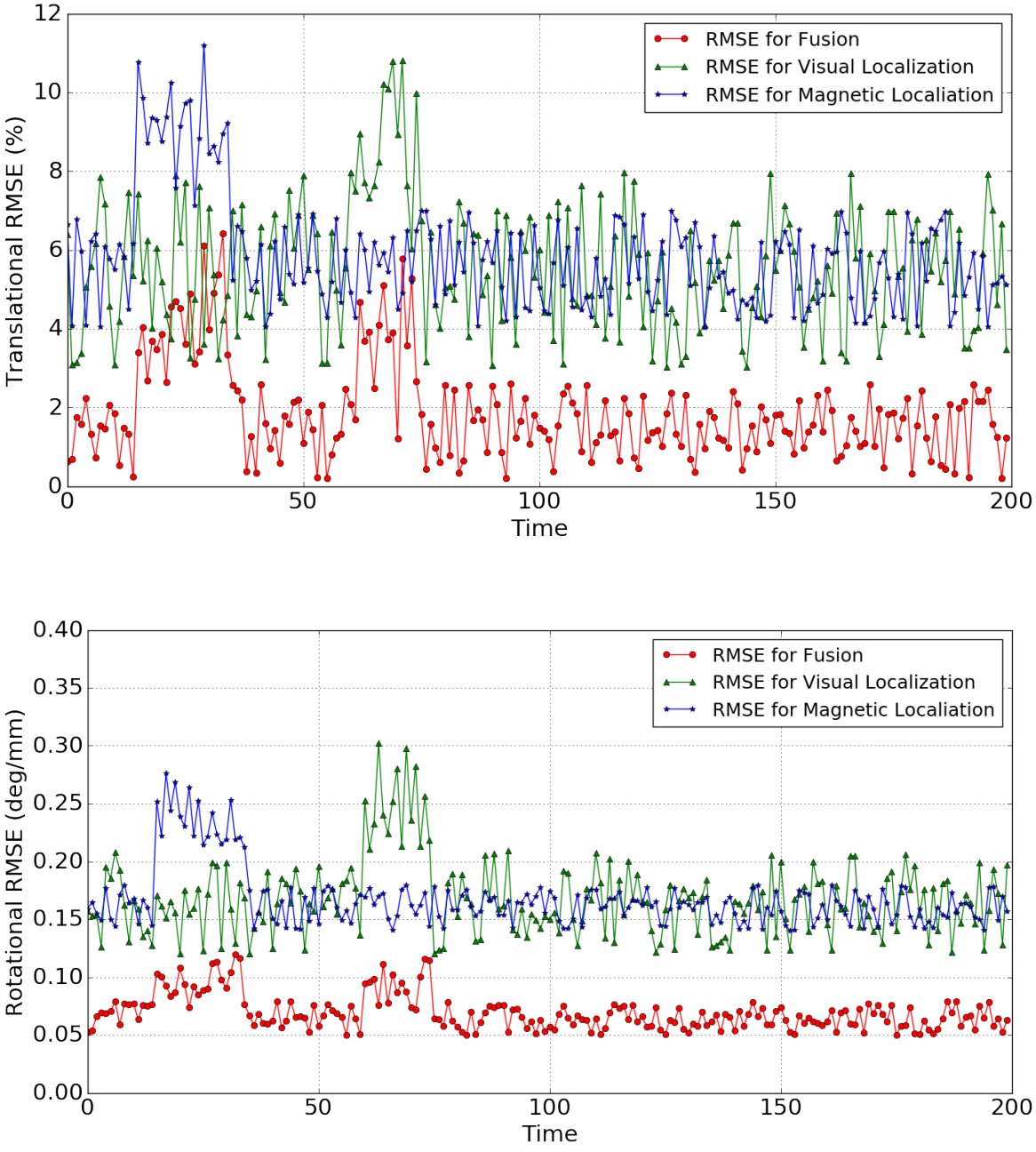


Figure 3.5. Translational (top) and rotational (bottom) RMSEs for multi-sensor fusion, visual localization and magnetic localization.

LSTM enabled learning motion dynamics across time sequences very effectively.

4. CONCLUSIONS

Several filtering methods have been proposed to obtain more accurate and computationally feasible sequential filters till date. The most popular, promising and novel filtering techniques for endoscopic capsule robots are discussed in the thesis. In the estimation problems, the Bayesian filters have been widely used since the first development of linear and Gaussian filters. The Kalman filter provides the optimal solution for the inference in dynamic linear-Gaussian systems with respect to MMSE. The optimal Kalman filter fails for the dynamical systems that has non-linear and non-Gaussian densities. Several extensions to the Kalman filter have been investigated to relax the linearity conditions of the filter such as EKF and UKF. In order to address the non-linear and non-Gaussian distributions, sequential Monte Carlo (SMC) based filters have been proposed as a promising solution, which comes with a computational cost. Thanks to the computational advances in the last decades, the SMC based filtering methods have been feasible, which are more robust than the Kalman filter methods. However, SMC based filters still requires the observation and the transition densities of the model to be defined. In many real world applications, the true densities may not have a mathematical mode, resulting in a lack of closed form solution. Specifically, linear motion assumptions of the traditional Kalman filter do not hold for endoscopic capsule robots. The non-linear motion model of the endoscopic capsule robot can be learned by a recurrent neural network from data. Following the success of the deep learning approaches in various domains in the recent years, these approaches have attracted attention in the estimation of the observation and transition models for the filtering algorithms. The LSTMs are able to capture both linear and non-linear motion of the endoscopic capsule robot without any assumptions about the type of motion of the robot. Furthermore, the LSTM incorporates a sliding window of the previous motion history during the learning to predict the next position of the robot thanks to the recurrent nature of the deep neural network. The LSTM is trained in a unique way to predict the next position of the mobile robot based on motion dynamics stored in the previous state of the robot. The LSTM estimation is input into a particle filter framework, which solves a critical challenge in the sensor fusion task predicting target

motion from noisy measurements delivered by the sensors attached to the robot.

In this study, we presented, to the best of our knowledge, the first particle filter based multi-sensor data fusion approach with a sensor failure detection and observation switching capability for endoscopic capsule robot localization. An LSTM architecture was used for non-linear motion model estimation of the capsule robot. The proposed system results in sub-millimetre scale precisions both for translational and rotational motions and outperforms both visual and magnetic sensors based localization techniques. As a future step, we consider to integrate a deep learning based noise-variance modelling functionality into our approach to eliminate sensor noise more efficiently. Moreover, in the kinematics modelling part of this work, the motion information is utilized to learn the dynamics, disregarding visual features of the endoscopic capsule robot. Features corresponding to motion characteristics of the robot can be incorporated from the endoscopic images, such as SURF, ORB features and features learned by a deep learning architecture, to improve measurement robot association.

REFERENCES

1. Son, D., M. D. Dogan and M. Sitti, "Magnetically actuated soft capsule endoscope for fine-needle aspiration biopsy", *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 1132–1139, IEEE, 2017.

2. Turan, M., Y. Almalioglu, H. Araujo, E. Konukoglu and M. Sitti, "A Non-Rigid Map Fusion-Based RGB-Depth SLAM Method for Endoscopic Capsule Robots", *arXiv preprint arXiv:1705.05444*, 2017.

3. Gers, F. A., J. Schmidhuber and F. Cummins, "Learning to Forget: Continual Prediction with LSTM", *Neural Computation*, Vol. 12, No. 10, pp. 2451–2471, 2000.

4. Turan, M., Y. Almalioglu, H. Araujo, E. Konukoglu and M. Sitti, "Deep EndoVO: A Recurrent Convolutional Neural Network (RCNN) based Visual Odometry Approach for Endoscopic Capsule Robots", *arXiv preprint arXiv:1708.06822*, 2017.

5. Turan, M., Y. Y. Pilavci, I. Ganiyusufoglu, H. Araujo, E. Konukoglu and M. Sitti, "Sparse-then-Dense Alignment based 3D Map Reconstruction Method for Endoscopic Capsule Robots", *arXiv preprint arXiv:1708.09740*, 2017.

6. Iddan, G., G. Meron, A. Glukhovsky and P. Swain, "Wireless capsule endoscopy", *Nature*, Vol. 405, No. 6785, pp. 417–417, 2000.

7. Sitti, M., H. Ceylan, W. Hu, J. Giltinan, M. Turan, S. Yim and E. Diller, "Biomedical applications of untethered mobile milli/microrobots", *Proceedings of the IEEE*, Vol. 103, No. 2, pp. 205–224, 2015.

8. Turan, M., Y. Almalioglu, E. Konukoglu and M. Sitti, "A Deep Learning Based 6 Degree-of-Freedom Localization Method for Endoscopic Capsule Robots", *arXiv preprint arXiv:1705.05435*, 2017.

9. Liao, Z., R. Gao, C. Xu and Z.-S. Li, "Indications and detection, completion, and retention rates of small-bowel capsule endoscopy: a systematic review", *Gastrointestinal endoscopy*, Vol. 71, No. 2, pp. 280–286, 2010.

10. Nakamura, T. and A. Terano, "Capsule endoscopy: past, present, and future", *Journal of gastroenterology*, Vol. 43, No. 2, pp. 93–99, 2008.

11. Pan, G. and L. Wang, "Swallowable wireless capsule endoscopy: progress and technical challenges", *Gastroenterology research and practice*, Vol. 2012, 2011.

12. Than, T. D., G. Alici, H. Zhou and W. Li, "A review of localization systems for robotic endoscopic capsules", *IEEE Transactions on Biomedical Engineering*, Vol. 59, No. 9, pp. 2387–2399, 2012.

13. Turan, M., A. Abdullah, R. Jamiruddin, H. Araujo, E. Konukoglu and M. Sitti, "Six Degree-of-Freedom Localization of Endoscopic Capsule Robots using Recurrent Neural Networks embedded into a Convolutional Neural Network", *arXiv preprint arXiv:1705.06196*, 2017.

14. Goenka, M. K., S. Majumder and U. Goenka, "Capsule endoscopy: Present status and future expectation", *World J Gastroenterol*, Vol. 20, No. 29, pp. 10024–10037, 2014.

15. Munoz, F., G. Alici and W. Li, "A review of drug delivery systems for capsule endoscopy", *Advanced drug delivery reviews*, Vol. 71, pp. 77–85, 2014.

16. Carpi, F., N. Kastelein, M. Talcott and C. Pappone, "Magnetically controllable gastrointestinal steering of video capsules", *IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 2, pp. 231–234, 2011.

17. Keller, H., A. Juloski, H. Kawano, M. Bechtold, A. Kimura, H. Takizawa and R. Kuth, "Method for navigation and control of a magnetically guided capsule endoscope in the human stomach", *RAS & EMBS International Conference on*

Biomedical Robotics and Biomechatronics (BioRob), pp. 859–865, IEEE, 2012.

18. Mahoney, A. W., S. E. Wright and J. J. Abbott, "Managing the attractive magnetic force between an untethered magnetically actuated tool and a rotating permanent magnet", *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 5366–5371, IEEE, 2013.

19. Turan, M., Y. Y. Pilavci, R. Jamiruddin, H. Araujo, E. Konukoglu and M. Sitti, "A fully dense and globally consistent 3D map reconstruction approach for GI tract to enhance therapeutic relevance of the endoscopic capsule robot", *arXiv preprint arXiv:1705.06524*, 2017.

20. Yim, S., E. Gultepe, D. H. Gracias and M. Sitti, "Biopsy using a magnetic capsule endoscope carrying, releasing, and retrieving untethered microgrippers", *IEEE Transactions on Biomedical Engineering*, Vol. 61, No. 2, pp. 513–521, 2014.

21. Petruska, A. J. and J. J. Abbott, "An omnidirectional electromagnet for remote manipulation", *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 822–827, IEEE, 2013.

22. Son, D., S. Yim and M. Sitti, "A 5-D localization method for a magnetically manipulated untethered robot using a 2-D array of Hall-effect sensors", *IEEE/ASME Transactions on Mechatronics*, Vol. 21, No. 2, pp. 708–716, 2016.

23. Bao, G., K. Pahlavan and L. Mi, "Hybrid localization of microrobotic endoscopic capsule inside small intestine by data fusion of vision and RF sensors", *IEEE Sensors Journal*, Vol. 15, No. 5, pp. 2669–2678, 2015.

24. Geng, Y. and K. Pahlavan, "Design, implementation, and fundamental limits of image and RF based wireless capsule endoscopy hybrid localization", *IEEE Transactions on Mobile Computing*, Vol. 15, No. 8, pp. 1951–1964, 2016.

25. Umay, I. and B. Fidan, "Adaptive magnetic sensing based wireless capsule localiza-

tion", *Medical Information and Communication Technology (ISMICT), 2016 10th International Symposium on*, pp. 1–5, IEEE, 2016.

26. Umay, I. and B. Fidan, "Adaptive wireless biomedical capsule tracking based on magnetic sensing", *International Journal of Wireless Information Networks*, Vol. 24, No. 2, pp. 189–199, 2017.

27. Gumprecht, J. D., T. C. Lueth and M. B. Khamesee, "Navigation of a robotic capsule endoscope with a novel ultrasound tracking system", *Microsystem technologies*, Vol. 19, No. 9-10, pp. 1415–1423, 2013.

28. Murphy, K. P., *Machine learning: a probabilistic perspective*, MIT press, 2012.

29. Bishop, C. M., *Pattern recognition and machine learning*, Springer, 2006.

30. Costa, O. L. V., M. D. Fragoso and R. P. Marques, *Discrete-time Markov jump linear systems*, Springer Science & Business Media, 2006.

31. Driessen, H. and Y. Boers, "An efficient particle filter for jump Markov nonlinear systems", *Target Tracking 2004: Algorithms and Applications, IEE*, pp. 19–22, IET, 2004.

32. Mazor, E., A. Averbuch, Y. Bar-Shalom and J. Dayan, "Interacting multiple model methods in target tracking: a survey", *IEEE Transactions on aerospace and electronic systems*, Vol. 34, No. 1, pp. 103–123, 1998.

33. Del Moral, P., A. Doucet and A. Jasra, "Sequential monte carlo samplers", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 68, No. 3, pp. 411–436, 2006.

34. Robert, C. and G. Casella, "A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data", *Statistical Science*, pp. 102–115, 2011.

35. Fitzgerald, W., "Markov chain Monte Carlo methods with applications to signal

processing", *Signal Processing*, Vol. 81, No. 1, pp. 3–18, 2001.

36. Einicke, G. A., *Smoothing, filtering and prediction*, InTech, 2012.

37. Julier, S. J., J. K. Uhlmann and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems", *American Control Conference, Proceedings of the 1995*, Vol. 3, pp. 1628–1632, IEEE, 1995.

38. Wan, E. A. and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation", *Adaptive Systems for Signal Processing, Communications, and Control Symposium. AS-SPCC.*, pp. 153–158, IEEE, 2000.

39. Berzuini, C., N. G. Best, W. R. Gilks and C. Larizza, "Dynamic conditional independence models and Markov chain Monte Carlo methods", *Journal of the American Statistical Association*, Vol. 92, No. 440, pp. 1403–1412, 1997.

40. Kitagawa, G., "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *Journal of computational and graphical statistics*, Vol. 5, No. 1, pp. 1–25, 1996.

41. Cemgil, A. T., H. J. Kappen and D. Barber, "A generative model for music transcription", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 2, pp. 679–694, 2006.

42. Murphy, K. P., "Bayesian map learning in dynamic environments", *Advances in Neural Information Processing Systems*, pp. 1015–1021, 2000.

43. Doucet, A., S. Godsill and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and computing*, Vol. 10, No. 3, pp. 197–208, 2000.

44. Visentini-Scarzanella, M., D. Stoyanov and G.-Z. Yang, "Metric depth recovery from monocular images using shape-from-shading and specularities", *Image Pro-*

cessing (ICIP), 2012 19th IEEE International Conference on, pp. 25–28, IEEE, 2012.

45. Dai, A., M. Nießner, M. Zollhöfer, S. Izadi and C. Theobalt, "BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Reintegration", *ACM Transactions on Graphics (TOG)*, Vol. 36, No. 3, p. 24, 2017.

46. Whelan, T., S. Leutenegger, R. Salas-Moreno, B. Glocker and A. Davison, "ElasticFusion: Dense SLAM without a pose graph", Robotics: Science and Systems, 2015.

47. Newcombe, R. A., S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking", *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pp. 127–136, IEEE, 2011.

48. Daniilidis, K., "Hand-eye calibration using dual quaternions", *The International Journal of Robotics Research*, Vol. 18, No. 3, pp. 286–298, 1999.

49. Caron, F., M. Davy, E. Duflos and P. Vanheeghe, "Particle filtering for multisensor data fusion with switching observation models: Application to land vehicle positioning", *IEEE transactions on Signal Processing*, Vol. 55, No. 6, pp. 2703–2719, 2007.

50. Hue, C., J.-P. Le Cadre and P. Perez, "Sequential Monte Carlo methods for multiple target tracking and data fusion", *IEEE Transactions on signal processing*, Vol. 50, No. 2, pp. 309–325, 2002.

51. Gilks, W. R. and P. Wild, "Adaptive rejection sampling for Gibbs sampling", *Applied Statistics*, pp. 337–348, 1992.

52. Walch, F., C. Hazirbas, L. Leal-Taixé, T. Sattler, S. Hilsenbeck and D. Cremers, "Image-based localization with spatial LSTMs", *arXiv preprint arXiv:1611.07890*,

2016.

53. Kingma, D. and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

54. Srivastava, N., G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting.", *Journal of machine learning research*, Vol. 15, No. 1, pp. 1929–1958, 2014.