

# A Multi-Rate State Observer for Visual Tracking of Magnetic Micro-Agents using 2D Slow Medical Imaging Modalities

Mert Kaya\*, Alper Denasi\*, Stefano Scheggi, Erdem Agbahca,  
ChangKyu Yoon, David H. Gracias, and Sarthak Misra

**Abstract**—Minimally invasive surgery can benefit greatly from utilizing micro-agents. These miniaturized agents need to be clearly visualized and precisely controlled to ensure the success of the surgery. Since medical imaging modalities suffer from low acquisition rate, multi-rate sampling methods can be used to estimate the intersample states of micro-agents. Hence, the sampling rate of the controller can be virtually increased even if the position data is acquired using a slow medical imaging modality. This study presents multi-rate Luenberger and Kalman state estimators for visual tracking of micro-agents. The micro-agents are tracked using sum of squared differences and normalized cross correlation based visual tracking. Further, the outputs of the two methods are merged to minimize the tracking error and prevent tracking failures. During the experiments, the micro-agents with different geometrical shapes and sizes are imaged using a 2D ultrasound machine and a microscope, and manipulated using electromagnetic coils. The multi-rate state estimation accuracy is measured using a high speed camera. The precision of the tracking and multi-rate state estimation are verified experimentally under challenging conditions. For this purpose, an elliptical shaped magnetic micro-agent with a length of 48 pixels is used. Maximum absolute error in  $x$  and  $y$  axes are 2.273 and 2.432 pixels for an 8-fold increase of the sample rate (25 frames per second), respectively. During the experiments, it was observed that the micro-agents could be tracked more reliably using normalized cross correlation based visual tracking and intersample states could be estimated more accurately using Kalman state estimator. Experimental results show that the proposed method could be used to track micro-agents in medical imaging modalities and estimate system states at intermediate time instants in real-time.

## I. INTRODUCTION

The field of micro-robotics has been witnessing a significant progress thanks to the developments in micro-fabrication, sensing and actuation technologies. Minimally invasive surgery (MIS) is among one of its most prominent and promising application domains. Utilizing micro-agents

M. Kaya, A. Denasi, S. Scheggi, and S. Misra are affiliated with the Surgical Robotics Laboratory, Department of Biomechanical Engineering, University of Twente, The Netherlands. M. Kaya, A. Denasi and S. Misra are also affiliated with the Department of Biomedical Engineering, University of Groningen and University Medical Centre Groningen, The Netherlands. Erdem Agbahca is affiliated with Department of Computer Engineering, Faculty of Technology, Selcuk University, Turkey. ChangKyu Yoon and D.H. Gracias are with the Department of Materials Science and Engineering, The Johns Hopkins University, USA. D.H. Gracias is also affiliated with the Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, USA. This project (ROBOTAR) has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation programme (Grant Agreement #638428). We also acknowledge support in part from the National Institutes of Health under Award RO1EB017742. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

\*These authors contributed equally to this work.

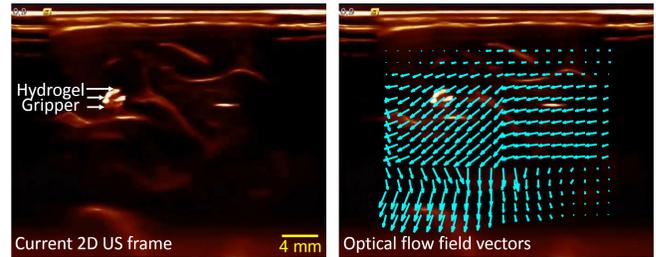


Fig. 1: Optical flow analysis of the current frame while a hydrogel gripper is magnetically steered. The image on the left shows the current 2D ultrasound (US) frame that contains the hydrogel gripper. The image on the right shows the optical flow field vectors overlaid on the current frame.

in MIS can reduce patient treatment and recovery time greatly compared to conventional surgery [1]. A successful medical intervention by an operator using micro-robotic agents requires clear visualization, accurate sensing and precise feedback control. Numerous studies can be found in the literature about micro-agents controlled using visual feedback. In the following, examples of some of the studies are outlined.

Scheggi *et al.* [2] investigated the closed-loop control of hydrogel grippers using electromagnetic actuation and 2D US images for feedback. Further, collision-free paths were planned using a combination of linear quadratic regulator and rapidly-exploring random tree algorithms. Keuning *et al.* [3] controlled paramagnetic micro-particles using 2D microscope images. Hong *et al.* [4] extracted the 3D position of microrobots from 2D microscope images based on digital inline holography. Further, a cylindrical permanent magnet was tracked and controlled in 3D. Martel *et al.* [5] developed a platform for steering of magnetic nanorobots using magnetic resonance imaging. In order to provide feedback to the controller, nanorobots were tracked in 3D using their magnetic signatures. Diller *et al.* [6] controlled multiple magnetic microrobots in 3D using magnetic gradient. Microrobots were tracked using two CCD cameras.

The aforementioned studies utilize medical imaging modalities as a feedback source for the controllers since traditional sensors like encoders cannot be embedded on the micro-agents due to their small size. But, these imaging modalities suffer from low acquisition rate. Image acquisition rates of magnetic resonance imaging and computed tomography are in the range of 10 to 20 frames per seconds (*fps*) and 6 to 15 *fps*, respectively. Maximum image rates of fluoroscopy and ultrasound imaging can reach up to 30

*fps* and *100 fps*, respectively [1], [7]. An appealing remedy for this drawback is multi-rate state estimation. This method provides estimates of the system states for intermediate time instants where measurements are not available. Multi-rate state estimation has been applied to various fields such as vehicle motion control systems [8], hard disk drive servo control systems [9], and visual servoing of manipulators [10]. In this study, we propose to apply multi-rate Luenberger and Kalman state estimators for the visual tracking outputs of micro-agents in order to have intermediate position data between two consecutive frames. The contribution of this study is that the multi-rate state estimation is applied to the field of micro-robotics to virtually increase the sampling rate of position data for control using low-rate visual feedback.

In the literature, detection-based tracking methods are commonly used to find the micro-agents location in the image plane [2], [3]. In these methods, micro-agents are first detected in each frame and then Kalman filter is applied for tracking. Detection of the micro-agents in each frame requires thresholding procedure and values for thresholding change by image. Thus, tracking is prone to failure unless the thresholding values are tuned properly. Besides, this approach ignores the motion information of the micro-agents in the image plane. But, motion information is an important cue and can be used to track the micro-agents. Due to the shortcomings of this approach, micro-agents are tracked in the image plane using similarity functions also known as template-based tracking. In this method, an object is detected once and then tracked by calculating its motion parameters between two close frames without applying any specific detection method. The pixels of micro-agents have considerable motion information when they are steered. The motion analysis of an electromagnetically actuated hydrogel gripper with images acquired from 2D US is shown in Fig. 1. The motion in this image is analyzed using the Lukas-Kanade optical flow method [11]. The analyzed motion is represented as a vector field and overlaid onto the current image that is shown in Fig. 1. This example shows that micro-agents with different shapes can often be visually tracked using similarity functions with images acquired from any medical imaging modality. In this study, the sum of squared differences (SSD) and the normalized cross correlation (NCC) cost functions are used as similarity functions. During the tracking, the outputs of SSD and NCC methods are merged using Luenberger and Kalman state observers to minimize the tracking error.

In the following section, visual tracking of the micro-agents is explained in detail. Section III states the multi-rate state estimation problem. This is followed by observer based fusion of SSD and NCC visual tracking methods in Section IV. In Section V, experimental setup is introduced and results are given. Finally, conclusions are drawn in Section VI.

## II. VISUAL TRACKING OF MAGNETIC MICRO-AGENTS

Visual tracking provides an estimate of the position of an object in the image plane over time by registering the previous appearance of the object with the current one. The previous image is called the template  $T(\mathbf{x})$ , and the recent

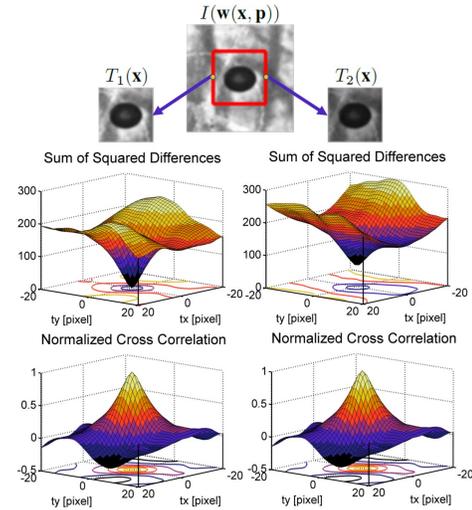


Fig. 2: 3D similarity plots of sum of squared differences and normalized cross correlation cost functions. Left columns contains similarity plots between  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$  and  $T_1(\mathbf{x})$ ; right column contains similarity plots between  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$  and  $T_2(\mathbf{x})$ . Further,  $tx$  and  $ty$  are the pixel increments of  $T_1(\mathbf{x})$  and  $T_2(\mathbf{x})$  with respect to the center of  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$ .

one is called the current image  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$ , where  $\mathbf{w}(\mathbf{x}, \mathbf{p})$  denotes the motion model (also known as warping function). This model maps the vector  $\mathbf{x} = [x, y]$  with the  $x$ - and  $y$ - coordinates of  $T(\mathbf{x})$  onto  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$  using the motion parameter vector,  $\mathbf{p}$ . The main goal of registration is to find the vector  $\mathbf{p}$  that minimizes the difference between  $T(\mathbf{x})$  and  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$  using an optimization algorithm. Briefly, registration process consists of two consecutive stages performed iteratively. In the first stage, the similarity between  $T(\mathbf{x})$  and  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$  is measured using a similarity function. In the second stage, the vector  $\mathbf{p}$  is calculated.

SSD and NCC cost functions are used to measure the similarity during the registration process. In order to evaluate the performance of these functions, a template image  $T_1(\mathbf{x})$  is cropped from  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$  that contains the micro-agent. The similarity between this image pair is measured by translating  $T_1(\mathbf{x})$  from  $-20$  to  $20$  pixels with  $1$  pixel increment in  $x$ - and  $y$ - axes from the center of  $I(\mathbf{w}(\mathbf{x}, \mathbf{p}))$ . A second template  $T_2(\mathbf{x})$  is also created by changing the contrast of  $T_1(\mathbf{x})$  to evaluate the performance of the cost functions under intensity changes. Similarity values are plotted in Fig. 2. As seen from Fig. 2, SSD values reach local minimum and NCC values reach local maximum for two cases in the center of the plots, which confirms that SSD and NCC cost functions can be used for visual tracking of micro-agents. In the rest of this section, motion model selection and visual tracking using SSD and NCC similarity functions are explained.

### A. Motion Model

One of the most significant layers of visual tracking is determining the motion of a micro-agent in the image plane. The motion model describes the transformation between the template and current images. The motion of a micro-agent in the image plane can be decomposed into translation, rotation, and scaling transformations. These transformations can be

described by an affine motion model,  $\mathbf{w}(\mathbf{x}, \mathbf{p}) : \mathbb{R}^2 \times \mathbb{R}^6 \rightarrow \mathbb{R}^2$ , given as follows:

$$\mathbf{w}(\mathbf{x}, \mathbf{p}) = \begin{bmatrix} (1+p_1) & p_3 \\ p_2 & (1+p_4) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} p_5 \\ p_6 \end{bmatrix}, \quad (1)$$

where  $\mathbf{p} \in \mathbb{R}^6$  is the 6-DOF motion parameter vector. Further, the motion model (1) can be used to track micro-agents with different geometric shapes.

During the visual tracking of micro-agents, illumination variations occur in microscope images and intensity variations occur in other medical imaging modalities (e.g. ultrasound). For successful tracking, illumination and intensity variations must be compensated. A scale-offset model can compensate these variations and consequently the compensated current image  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$ , is written as follows:

$$I^*(\mathbf{w}(\mathbf{x}, \mathbf{q})) = (\alpha + 1)I(\mathbf{w}(\mathbf{x}, \mathbf{p})) + \beta, \quad (2)$$

where  $(\alpha, \beta)$  are the model parameters. Using the affine motion and scale-offset models, a new parameter vector  $\mathbf{q} \in \mathbb{R}^8$  where  $\mathbf{q} = [\mathbf{p}^T, \alpha, \beta]^T$  are calculated for the transformation from the template to the current image during the visual tracking. The calculation of  $\mathbf{q}$  vector using SSD and NCC cost functions is explained next.

### B. SSD-based Visual Tracking

SSD between  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$  and  $T(\mathbf{x})$  can be computed as:

$$SSD(\mathbf{q}) = \sum_{\mathbf{x}} [I^*(\mathbf{w}(\mathbf{x}, \mathbf{q})) - T(\mathbf{x})]^2. \quad (3)$$

The main goal of SSD-based visual tracking is to find the vector  $\mathbf{q}$  that minimizes the SSD between  $T(\mathbf{x})$  and  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$ . For this purpose, forward or inverse compositional methods [12], first order optimization, or efficient second order minimization (ESM) methods [13] can be used. Forward and inverse compositional methods include calculation of Hessian matrix. However, calculation of Hessian matrix is computationally expensive. First order optimization method directly ignores calculation of Hessian matrix. But, Hessian matrix should be included in the optimization method to increase the precision of the tracking. In the ESM method, Hessian matrix is estimated using Jacobian of the current and template images, which makes this method more favorable within the four optimization methods. Using this estimation approach, tracking becomes more robust to noise. Besides, its convergence rate is higher as compared to the other optimization methods since the computation time of the  $\mathbf{q}$  vector decreases. Therefore, ESM method is used to find the vector  $\mathbf{q}$  in this study. Using ESM method,  $\mathbf{p}$  is iteratively computed by accumulating  $\Delta\mathbf{q}$  in each iteration ( $\mathbf{q} \leftarrow \mathbf{q} + \Delta\mathbf{q}$ ).  $\Delta\mathbf{q}$  is computed as follows:

$$\Delta\mathbf{q} = -2(\mathbf{J}(\mathbf{q}_0) + \mathbf{J}(\mathbf{q}_c))^\dagger (I^*(\mathbf{w}(\mathbf{x}, \mathbf{q})) - T(\mathbf{x})), \quad (4)$$

where  $\dagger$  is the operator for the Moore-Penrose pseudoinverse of a matrix. Further,  $\mathbf{J}(\mathbf{q}_0)$  and  $\mathbf{J}(\mathbf{q}_c)$  are the Jacobian of  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$  [14], [15]. Before the iteration loop starts,  $\mathbf{J}(\mathbf{q}_0)$  is pre-computed. During the iterations, first  $\mathbf{J}(\mathbf{q}_c)$  is computed and then  $\Delta\mathbf{q}$  is calculated using (4). Iterations last until the number of iterations reach a predefined maximum iteration number or sum of  $\Delta\mathbf{q}$  vector elements are smaller than the predefined threshold.

### C. NCC-based Visual Tracking

NCC between  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$  and  $T(\mathbf{x})$  can be computed as:

$$NCC(\mathbf{q}) = \frac{\sum_{\mathbf{x}} (I^*(\mathbf{w}(\mathbf{x}, \mathbf{q})) - \bar{I}^*)(T(\mathbf{x}) - \bar{T})}{\sqrt{\sum_{\mathbf{x}} (I^*(\mathbf{w}(\mathbf{x}, \mathbf{q})) - \bar{I}^*)^2} \sqrt{\sum_{\mathbf{x}} (T(\mathbf{x}) - \bar{T})^2}}, \quad (5)$$

where  $\bar{T}$  and  $\bar{I}^*$  are the mean values of  $T(\mathbf{x})$  and  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$ , respectively. The main goal of NCC-based visual tracking is to find the  $\mathbf{q}$  that maximizes the NCC between  $T(\mathbf{x})$  and  $I^*(\mathbf{w}(\mathbf{x}, \mathbf{q}))$ . Using Newton optimization method,  $\mathbf{q}$  can be iteratively computed by accumulating  $\Delta\mathbf{q}$  in each iteration ( $\mathbf{q} \leftarrow \mathbf{q} + \Delta\mathbf{q}$ ).  $\Delta\mathbf{q}$  can be computed with the inverse of Hessian ( $\mathbf{H}^{-1}$ ) matrix and gradient ( $\mathbf{g}$ ) vector of (5) as follows [14]:

$$\Delta\mathbf{q} = -\mathbf{H}^{-1}\mathbf{g}. \quad (6)$$

Similar to SSD-based visual tracking, iterations last until the sum of the elements of the vector  $\Delta\mathbf{q}$  is smaller than the predefined threshold or the number of iterations reaches the predefined maximum iteration number.

### D. Pyramidal Implementation

In our experiments, it was observed that the motion of the micro-sized agents can be very large in the image plane. In this situation, the displacement of the agents between the previous and the current frames can be large. Hence, the number of iterations to calculate the motion parameters between the template and current images increases dramatically. Further, the agents can be so far away from the search region in the image plane that tracking might fail. In order to increase the convergence rate and robustness of the tracking, pyramidal implementation of the proposed template-based tracking method is applied. During the tracking, optimization methods are first applied to the coarsest level of pyramid and then, results are transferred to the next finer level of the pyramid. The number of pyramid layer for the frames in our data set and imaging system was selected as two [11]. If the number of pyramid layers in the experiments was more than two, significant texture loss was observed because our maximum template size is  $60 \times 60$  square pixels.

### E. Template Update Strategy

During the tracking, the template images are updated with a drift correction strategy to minimize the registration error and prevent template drifts [16]. This strategy consists of two steps. In the first step, the template and current images are registered. In the second step, the output image obtained in the first step is registered with the master template. The master template is the first appearance of the micro-agent and updated every  $1.5 \times fps$  frames. After the second step, the template image is updated with registration output obtained in the second step and position of the micro-agents is obtained in the image. This position serves as a measurement to the multi-rate state estimators presented in the next section.

## III. MULTI-RATE STATE ESTIMATION

In this section, a simplified kinematics-based model used to design the multi-rate state estimators is described. This is followed by the theory and application of a multi-rate Luenberger state observer and a multi-rate Kalman filter.

### A. System Description

In the following, we assume that the dynamics of the micro-agents can be approximated by a linear and time-invariant model. Further, the influence of deviations of the actual system from the linear system are modeled with a disturbance term. It is possible to use a sophisticated nonlinear model for multi-rate state estimation. However, the problem with such a model is the identification of its parameters and further the validation of these parameters due to the unavailability of sensors or ground truth in micro-robotics. Hence, a disturbance term is used to locally represent the contributions of such nonlinear dynamics. The system (plant) dynamics is expressed in continuous-time as follows:

$$\dot{\mathbf{x}}_p = \mathbf{A}_{pc}\mathbf{x}_p + \mathbf{B}_{pc}\mathbf{u} + \mathbf{B}_{pc}\mathbf{d}, \quad (7)$$

$$\mathbf{y} = \mathbf{C}_{pc}\mathbf{x}_p + \mathbf{D}_{pc}\mathbf{u} + \mathbf{v}, \quad (8)$$

where  $\mathbf{x}_p \in \mathbb{R}^{n_p}$  and  $\mathbf{u} \in \mathbb{R}^m$  are the state of the plant and the control input, respectively. Further,  $\mathbf{d} \in \mathbb{R}^{n_d}$ ,  $\mathbf{v} \in \mathbb{R}^l$  and  $\boldsymbol{\eta} \in \mathbb{R}^l$  represent the disturbance, the measurement noise and the measured output, respectively. The measurement noise is assumed to be zero-mean, Gaussian and white. Further, the dynamics of the disturbance term ((7) and (8)) is denoted as:

$$\dot{\boldsymbol{\eta}}_d = \mathbf{A}_{dc}\boldsymbol{\eta}_d + \mathbf{B}_{dc}\mathbf{w}, \quad (9)$$

$$\mathbf{d} = \mathbf{C}_{dc}\boldsymbol{\eta}_d, \quad (10)$$

where  $\boldsymbol{\eta}_d \in \mathbb{R}^{n_\eta}$  and  $\mathbf{w} \in \mathbb{R}^{n_w}$  are the state of the disturbance dynamics and an external driving signal, respectively. The signal  $\mathbf{w}$  can either be deterministic but unknown or stochastic with the assumption of being zero-mean, Gaussian and white. If the signal  $\mathbf{w}$  is deterministic but unknown and bounded, then a Luenberger state observer can be designed. But, if the signal  $\mathbf{w}$  is stochastic with the assumption of being zero-mean, Gaussian and white, then a Kalman filter can be designed. The plant dynamics given by (7)-(8) and the disturbance dynamics given by (9)-(10) can be combined into the following augmented system:

$$\dot{\mathbf{x}} = \mathbf{A}_c\mathbf{x} + \mathbf{B}_c\mathbf{u} + \mathbf{B}_{wc}\mathbf{w}, \quad (11)$$

$$\mathbf{y} = \mathbf{C}_c\mathbf{x} + \mathbf{D}_c\mathbf{u} + \mathbf{v}, \quad (12)$$

where

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{A}_{pc} & \mathbf{B}_{pc}\mathbf{C}_{dc} \\ \mathbf{0}_{n_d \times n_p} & \mathbf{A}_{dc} \end{bmatrix}, \mathbf{B}_c = \begin{bmatrix} \mathbf{B}_{pc} \\ \mathbf{0}_{n_d \times m} \end{bmatrix}, \mathbf{B}_{wc} = \begin{bmatrix} \mathbf{0}_{n_p \times n_w} \\ \mathbf{B}_{dc} \end{bmatrix}, \quad (13)$$

$$\mathbf{C}_c = [\mathbf{C}_{pc} \quad \mathbf{0}_{l \times n_\eta}], \quad \mathbf{D}_c = \mathbf{D}_{pc}, \quad \mathbf{x} = [\mathbf{x}_p^T \quad \boldsymbol{\eta}_d^T]^T,$$

where  $\mathbf{0}$  is a zero-matrix of appropriate size. For systems with a low sampling rate, an appropriate discretization method should be selected. The sampled representation of a linear time-invariant system given by (11)-(12) can be obtained exactly using the exact discretization method. The system is discretized with a fast model sampling time,  $T_f \in \mathbb{R}_{>0}$  and the measured outputs are obtained with a slower sampling time,  $T_s \in \mathbb{R}_{>0}$ . These two sampling times satisfy  $T_s = NT_f$  where  $N \in \mathbb{Z}_{>0}$ . In imaging systems,  $T_s$  corresponds to reciprocal of the frames per second (1/fps) of the system. The block diagram of the dynamic system ((11) and (12)) with different sampling rates is shown in Fig. 3.

After discretizing (11) and (12), the resulting system of difference equations are given as follows:

$$\mathbf{x}[(i+1)T_f] = \mathbf{A}_d\mathbf{x}[iT_f] + \mathbf{B}_d\mathbf{u}[iT_f] + \mathbf{B}_{wd}\mathbf{w}[iT_f], \quad (14)$$

$$\mathbf{y}[jT_s] = \mathbf{C}_d\mathbf{x}[jT_s] + \mathbf{D}_d\mathbf{u}[jT_s] + \mathbf{v}[jT_s]. \quad (15)$$

The matrices in (14)-(15) can be computed using

$$\begin{bmatrix} \mathbf{A}_d & \mathbf{B}_d & \mathbf{B}_{wd} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} = \exp \left( \begin{bmatrix} \mathbf{A}_c & \mathbf{B}_c & \mathbf{B}_{wc} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} T_f \right) \quad (16)$$

$$\mathbf{C}_d = \mathbf{C}_c, \quad \mathbf{D}_d = \mathbf{D}_c \quad (17)$$

where  $\exp(\cdot)$  is the matrix exponential operator [17]. Here,  $\mathbf{0}$  and  $\mathbf{I}$  are the zero and identity matrices of appropriate size, respectively. For clarity of notation, the sampling time variables in (14)-(15) are omitted, and we obtain

$$\mathbf{x}(i+1) = \mathbf{A}_d\mathbf{x}(i) + \mathbf{B}_d\mathbf{u}(i) + \mathbf{B}_{wd}\mathbf{w}(i), \quad (i = 1, 2, 3, \dots), \quad (18)$$

$$\mathbf{y}(j) = \mathbf{C}_d\mathbf{x}(j) + \mathbf{D}_d\mathbf{u}(j) + \mathbf{v}(j), \quad (j = N, 2N, \dots). \quad (19)$$

The system (18)-(19) can be rewritten using a unified time step  $k$  ( $k = 1, 2, 3, \dots$ ) as follows:

$$\mathbf{x}(kN + c + 1) = \mathbf{A}_d\mathbf{x}(kN + c) + \mathbf{B}_d\mathbf{u}(kN + c) + \mathbf{B}_{wd}\mathbf{w}(kN + c),$$

$$\mathbf{y}(kN) = \mathbf{C}_d\mathbf{x}(kN) + \mathbf{D}_d\mathbf{u}(kN) + \mathbf{v}(kN), \quad (20)$$

where  $c = 0, 1, \dots, N-1$ . The multi-rate system of (20) can be lifted into a single-rate system ( $T_s$ ) as follows:

$$\mathbf{x}(kN + N) = \mathbf{A}_e\mathbf{x}(kN) + \mathbf{u}_e + \mathbf{w}_e, \quad (21)$$

$$\mathbf{y}(kN) = \mathbf{C}_d\mathbf{x}(kN) + \mathbf{D}_d\mathbf{u}(kN) + \mathbf{v}(kN), \quad (22)$$

where

$$\mathbf{A}_e = \mathbf{A}_d^N, \quad \mathbf{u}_e = \sum_{c=0}^{N-1} \mathbf{A}_d^{N-1-c} \mathbf{B}_d \mathbf{u}(kN + c), \quad (23)$$

$$\mathbf{w}_e = \sum_{c=0}^{N-1} \mathbf{A}_d^{N-1-c} \mathbf{B}_{wd} \mathbf{w}(kN + c). \quad (24)$$

In the following, as an example of (11)-(12), the model used for estimating the 2D positions of a micro-agent in camera or 2D ultrasound images is explained. Consider the x- and y- position coordinates of the micro-agent denoted by  $p_x \in \mathbb{R}$ ,  $p_y \in \mathbb{R}$  and the corresponding velocities denoted by  $v_x \in \mathbb{R}$ ,  $v_y \in \mathbb{R}$ . We consider the case when there are no control inputs, thus  $\mathbf{u} = \mathbf{0}$  holds. Considering the state vector  $\mathbf{x}_p = [p_x \quad p_y \quad v_x \quad v_y]^T$  and the disturbance vector  $\mathbf{d} = [d_x \quad d_y]^T$ , the system dynamics is given as follows:

$$\underbrace{\begin{bmatrix} \dot{p}_x \\ \dot{p}_y \\ \dot{v}_x \\ \dot{v}_y \end{bmatrix}}_{\dot{\mathbf{x}}_p} = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{A}_{pc}} \underbrace{\begin{bmatrix} p_x \\ p_y \\ v_x \\ v_y \end{bmatrix}}_{\mathbf{x}_p} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{B}_{pc}} \underbrace{\begin{bmatrix} d_x \\ d_y \end{bmatrix}}_{\mathbf{d}}, \quad (25)$$

$$\mathbf{y} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{C}_{pc}} \mathbf{x}_p + \underbrace{\begin{bmatrix} v_x \\ v_y \end{bmatrix}}_{\mathbf{v}}, \quad (26)$$

where the matrices given by (7)-(8) are used. The system dynamics described by  $\mathbf{A}_{pc}$  in (25) with  $\mathbf{d} = \mathbf{0}$  is commonly

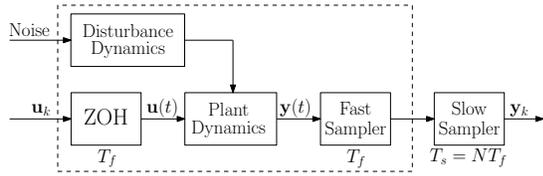


Fig. 3: The block diagram of a dynamic system sampled at different rates is presented. The system is discretized with a fast model sampling time  $T_f$  and the measured outputs are obtained with a slower sampling time  $T_s$ .

used in vision applications in the literature. However, in order to have a more general yet simple model, a disturbance term should be included. There are different ways to model the disturbance term depending on the application. In order to keep the formulation sufficiently general a polynomial function is selected [18]. The disturbance term can be locally represented by an  $(n-1)^{\text{th}}$  degree family of Taylor polynomial function of time as follows:

$$d_j(t) = \sum_{i=0}^{n-1} d_{j,i} t^i + d_{j,r}(t), \quad (27)$$

where  $j \in \{x, y\}$ ,  $d_{j,i} \in \mathbb{R}$  and  $d_{j,r}(t) \in \mathbb{R}$  are the coefficients of the polynomial and a residual term, respectively. It is assumed that the residual term,  $d_{j,r}(t)$ , is such that its time derivatives for  $i \geq n$  satisfy  $|d_{j,r}^{(i)}(t)| \leq \gamma_{i-r} \approx 0$ , thus they are all uniformly absolutely bounded and small enough to be negligible [18]. This implies that the residual term is slowly varying with respect to time. When, the degree of the polynomial is  $n$  we obtain,

$$\underbrace{\begin{bmatrix} \hat{\eta}_{d,1} \\ \hat{\eta}_{d,2} \\ \vdots \\ \hat{\eta}_{d,n-1} \\ \hat{\eta}_{d,n} \end{bmatrix}}_{\hat{\eta}_d} = \underbrace{\begin{bmatrix} \mathbf{0}_2 & \mathbf{I}_2 & \dots & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \dots & \mathbf{0}_2 & \mathbf{0}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_2 & \mathbf{0}_2 & \dots & \mathbf{0}_2 & \mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \dots & \mathbf{0}_2 & \mathbf{0}_2 \end{bmatrix}}_{\mathbf{A}_{dc}} \underbrace{\begin{bmatrix} \eta_{d,1} \\ \eta_{d,2} \\ \vdots \\ \eta_{d,n-1} \\ \eta_{d,n} \end{bmatrix}}_{\eta_d} + \underbrace{\begin{bmatrix} \mathbf{0}_2 \\ \mathbf{0}_2 \\ \vdots \\ \mathbf{0}_2 \\ \mathbf{I}_2 \end{bmatrix}}_{\mathbf{B}_{dc}} \underbrace{\begin{bmatrix} d_{x,r}^{(n)} \\ d_{y,r}^{(n)} \\ \vdots \end{bmatrix}}_{\mathbf{w}}, \quad (28)$$

$$\underbrace{\begin{bmatrix} d_x & d_y \end{bmatrix}}_{\mathbf{d}}^T = \underbrace{\begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 & \dots & \mathbf{0}_2 & \mathbf{0}_2 \end{bmatrix}}_{\mathbf{C}_{dc}} \eta_d,$$

where  $\mathbf{0}_2$  and  $\mathbf{I}_2$  are zero and identity matrices, respectively. Further, in (28) each element of the state vector  $\eta_d$  satisfies  $\eta_{d,i} \in \mathbb{R}^2$ . In the following subsections, the applications of multi-rate Luenberger state observer and multi-rate Kalman filter to the system described by (21) and (22) are detailed.

### B. A Multi-rate Luenberger Observer

The standard Luenberger observer uses the model of a given system together with a suitable update term to estimate state variables that cannot be measured. In order to apply it to the multi-rate augmented system dynamics ((14)-(15)) certain modifications are required [19]. First of all the system (21)-(22) should be observable. This condition is satisfied for the micro-agent dynamics described in Section III-A. This is satisfied if the observability matrix  $\mathcal{O}^T = \begin{bmatrix} \mathbf{C}_d^T & \mathbf{A}_e^T \mathbf{C}_d^T & \dots & (\mathbf{A}_e^{n_p+n_d-1})^T \mathbf{C}_d^T \end{bmatrix}$  has full rank, i.e.  $\text{rank}(\mathcal{O}) = n_p + n_d$ . Since  $\mathbf{y}(kN)$  is only available every  $N^{\text{th}}$  step for the samples between  $kN$  and  $kN + N$ , the value  $\mathbf{y}(kN)$  has to be used in the feedback term. Thus, the

observer is given as follows:

$$\hat{\mathbf{x}}(kN + 1 + c) = \mathbf{A}_d \hat{\mathbf{x}}(kN + c) + \mathbf{B}_d \mathbf{u}(kN + c) - \mathbf{L}(\hat{\mathbf{y}}(kN) - \mathbf{y}(kN)) \quad (29)$$

for  $c = 0, 1, \dots, N-1$  with the observer state  $\hat{\mathbf{x}}(kN + c)$  and feedback gain matrix  $\mathbf{L}$ . The observer state  $\hat{\mathbf{x}}(kN + N)$  after  $N$  time steps is given as follows:

$$\hat{\mathbf{x}}(kN + N) = \mathbf{A}_e \hat{\mathbf{x}}(kN) + \mathbf{u}_e - \left( \sum_{c=0}^{N-1} \mathbf{A}_d^c \right) \mathbf{L}(\hat{\mathbf{y}}(kN) - \mathbf{y}(kN)) \quad (30)$$

and the observed output  $\hat{\mathbf{y}}(kN)$  is given by

$$\hat{\mathbf{y}}(kN) = \mathbf{C}_d \hat{\mathbf{x}}(kN) + \mathbf{D}_d \mathbf{u}(kN). \quad (31)$$

Using (21), (22) and (30) and defining the observation error as  $\mathbf{e}(kN) = \hat{\mathbf{x}}(kN) - \mathbf{x}(kN)$ , the error dynamics can be given as follows:

$$\mathbf{e}(kN + N) = (\mathbf{A}_e - \bar{\mathbf{L}} \mathbf{C}_d) \mathbf{e}(kN) + \bar{\mathbf{L}} \mathbf{v}(kN) - \mathbf{w}_e, \quad (32)$$

where  $\bar{\mathbf{L}} = (\sum_{c=0}^{N-1} \mathbf{A}_d^c) \mathbf{L}$  is an auxiliary feedback gain matrix. This is done such that if the system is observable, the eigenvalues of the nominal system (i.e. for  $\mathbf{v} = 0$  and  $\mathbf{w}_e = 0$ ) can be placed at arbitrary locations, for instance using Ackermann's formula. Consequently, the observer gain matrix  $\mathbf{L}$  can be selected as follows:

$$\mathbf{L} = \left( \sum_{c=0}^{N-1} \mathbf{A}_d^c \right)^{-1} \bar{\mathbf{L}} \quad (33)$$

when the inverse of the matrix  $\sum_{c=0}^{N-1} \mathbf{A}_d^c$  exists. The nominal system is asymptotically stable if all of the eigenvalues are inside the unit disk. The stability of the error dynamics (32) can be shown using a variation of the input-to-state stability theorem [20].

### C. A Multi-rate Kalman Filter

The application of a Kalman Filter to a system with multiple sampling rates where the measurement sampling rate is low, is different as compared to a standard Kalman filter [9]. The multi-rate Kalman filter involves two steps; prediction and correction. In the prediction step, there are no available measurements in between the measurement sampling instants. Therefore, the prediction is based on the augmented system dynamics (18)-(19) which is an open-loop system. Denote  $\hat{\mathbf{x}}(i|j)$  as the estimate of  $\hat{\mathbf{x}}(i)$  based on the measurement before and on time  $jT_s$ . At the time instants  $t = (Nk + m)T_f$  ( $\forall m = 1, 2, \dots, N$ ), the fast-rate estimate is given as follows:

$$\hat{\mathbf{x}}(Nk + m|Nk) = \mathbf{A}_d^m \hat{\mathbf{x}}(Nk|Nk) + \sum_{c=0}^{m-1} \mathbf{A}_d^{m-1-c} \mathbf{B}_d \mathbf{u}(kN + c). \quad (34)$$

At the time instants  $t = N(k+1)T_f$ , new measurements  $\mathbf{y}(Nk + N)$  are available. Therefore, in the correction step the estimate is given as follows:

$$\hat{\mathbf{x}}(Nk + N|Nk + N) = \hat{\mathbf{x}}(Nk + N|Nk) + \mathbf{K}(k+1) [\mathbf{y}(Nk + N) - \mathbf{C}_d \hat{\mathbf{x}}(Nk + N|Nk)]. \quad (35)$$

Consequently, the gain matrix  $\mathbf{K}$  is updated through

$$\mathbf{K}(k+1) = \mathbf{M}(k+1) \mathbf{C}_d^T [\mathbf{C}_d \mathbf{M}(k+1) \mathbf{C}_d^T + \mathbf{V}]^{-1} \quad (36)$$

$$\begin{aligned} \mathbf{M}(k+1) &= \mathbf{A}_e \mathbf{M}(k) \mathbf{A}_e^T + \mathbf{W}_e \\ &- \mathbf{A}_e \mathbf{M}(k) \mathbf{C}_d^T [\mathbf{C}_d \mathbf{M}(k) \mathbf{C}_d^T + \mathbf{V}]^{-1} \mathbf{C}_d \mathbf{M}(k) \mathbf{A}_e^T, \end{aligned} \quad (37)$$

where  $\mathbf{V}$  and  $\mathbf{W}_e$  are the covariances of the measurement  $\mathbf{v}$  and process noise  $\mathbf{w}_e$ , respectively. The covariance for the measurement noise is given by  $\mathbf{V} = \text{cov}(\mathbf{v})$ . Since, the correction step occurs every  $N$  samples with the slow sampling time  $T_s$ , the covariance of the process noise for (21) should be used. The covariance of the process noise  $\mathbf{w}_e$  for the system (21) with the slow sampling time  $T_s$  can be related to the covariance of the process noise  $\mathbf{w}$  for the system (18) with the fast sampling time  $T_f$  using (24) as:

$$\begin{aligned} \mathbf{W}_e &= \text{cov} \left( \sum_{c=0}^{N-1} \mathbf{A}_d^{N-1-c} \mathbf{B}_{wd} \mathbf{w}(kN+c) \right) = E[\mathbf{w}_e \mathbf{w}_e^T] \\ &= \left( \sum_{c=0}^{N-1} \mathbf{A}_d^{N-1-c} \mathbf{B}_{wd} \right) \mathbf{W} \left( \sum_{c=0}^{N-1} \mathbf{A}_d^{N-1-c} \mathbf{B}_{wd} \right)^T, \end{aligned} \quad (38)$$

where  $\mathbf{W} = \text{cov}(\mathbf{w})$  and  $E[\cdot]$  is the expectation operator.

#### IV. COLLABORATIVE TRACKING USING OBSERVERS

One of the most important issues in visual tracking is error minimization. After the template and current images are registered to find the position of a micro-agent in the current image, there is a residual error called the registration error. If the registration error is not minimized during the tracking, the error accumulates over time and tracking eventually fails. In order to minimize the error, template update with drift correction strategy is used. This method is efficient when a micro-agent is tracked using only one of the visual tracking methods. In this study, SSD and NCC based visual tracking methods are used to collaboratively track the micro-agents [21]. If these two methods are used to track micro-agents in parallel, the outputs of these methods can be merged using either a Kalman or a Luenberger state observer to minimize the error. Thus, tracking becomes more accurate and resilient to failures. The data obtained from the SSD and NCC based visual tracking methods can be merged as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{SSD} \\ \mathbf{y}_{NCC} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_d \\ \mathbf{C}_d \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{v}_{SSD} \\ \mathbf{v}_{NCC} \end{bmatrix}, \quad (39)$$

where  $\mathbf{y}_{SSD}$  and  $\mathbf{y}_{NCC}$  are the outputs of the SSD and NCC visual tracking methods, respectively. Further,  $\mathbf{v}_{SSD}$  and  $\mathbf{v}_{NCC}$  denote the registration errors of SSD and NCC, respectively. The scheme of the merging is shown in Fig. 4.

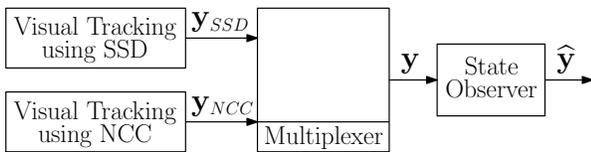


Fig. 4: Observer based merging scheme for sum of squared differences (SSD) and normalized cross correlation (NCC) based tracking.

#### V. EXPERIMENTS

In this section, the experimental setup is first described. This is followed by execution times and experimental results.

##### A. Experimental Setup

In the experiments, four iron-core coils placed on the same plane were used for planar manipulation of the micro-agents. Each coil was driven by an Elmo 1/60 Whistle DC servo drive. Further details about the electromagnetic setup can be found in our previous work [2], [22]. During the manipulation of the micro-agents, 2D US images were acquired using a Siemens ACUSON S2000 US machine with a 18L6HD linear 2D US probe. The acquired 2D US images were transferred to the computer using an Epiphan DVI2USB 3.0 frame grabber. Microscope images of the micro-agents were captured using an Optem FUSION 7:1 zoom lens module and a XIMEA xiQ USB 3.0 high speed camera pair.

##### B. Execution Times

The proposed method was implemented using both MATLAB and C++ with OpenCV library. All of the experiments were performed on a MacBook Pro which has a 2.9 GHz Intel Core i5 CPU with 8 GB RAM. During the tracking, the size of the template images was  $60 \times 60$  square pixels. The maximum iteration number and the predefined threshold for terminating the iterations were set to 20 and 0.01, respectively. The average execution times were measured as 69.68 [ms] in MATLAB and 13.22 [ms] in C++, respectively. The execution time of the proposed method was significantly reduced in C++ implementation, up to five fold, that allows the real-time execution of the proposed method with the frame rate of the imaging device, which is about 75 *fps*.

##### C. Experimental Results

1) *Visual Tracking Results*: During the experiments, magnetic hydrogel grippers, spherical and elliptical shaped micro-agents are used. All of the agents are imaged using a microscope and a 2D US machine. In total, 120652 images are acquired. During the image acquisition, minimum and maximum *fps* are 14 and 200, respectively. In order to evaluate the tracking performance, an elliptical shaped micro-agent was manipulated using the electromagnetic system for 65 seconds. During the manipulation, frames were acquired at 198 *fps* and 12801 frames were captured in total. The tracking performance was quantitatively evaluated by computing both NCC and a combination of Forward-Backward error and NCC (FB+NCC) values [23]. NCC value was computed between the template image and the image obtained after registration. In order to compute the FB+NCC value, the image obtained after registration, called the forward image, and the previous frame was registered to obtain the backward image. NCC+FB value was calculated by computing NCC value between the backward and the template images. NCC and FB+NCC values were computed as  $0.9943 \pm 0.0117$  and  $0.9970 \pm 0.0082$ , respectively. As mentioned in Section II, SSD and NCC based visual tracking methods can fail if the agent moves too much between the measurements. This was experimentally tested by magnetically steering an elliptical shaped micro-agent with a length of 48 pixels. It was observed that if the Euclidean distance between the agent positions in two consecutive frames is more than 38 pixels, which corresponds to a velocity of 163.57 mm/s, tracking fails. But such a movement is an extreme case in our system.

Visual tracking results of a magnetic hydrogel gripper with a tip-to-tip distance of 4 mm in 2D US images are shown in Fig. 5. The images are acquired at 30 fps and the gripper is manipulated using magnetic fields in a water filled tube. Synthetic artifacts are also added to the images to show that the proposed visual tracking method can track the micro-agents in a dynamic background and under realistic conditions. In order to obtain synthetic artifacts, abdominal region of a volunteer is scanned and recorded with a 2D US probe. After the raw image is acquired during manipulation of the gripper, the raw and abdominal images are superimposed.

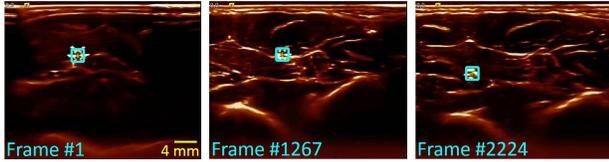


Fig. 5: Tracking results for a hydrogel gripper with a tip-to-tip distance of 4 mm in 2D ultrasound images. Please refer to the accompanying video.

A 1 mm elliptical shaped micro-agent is tracked in microscope images. Before image acquisition, a printed vein pattern is placed at the bottom of the Petri dish and the agent is magnetically steered. The motion of the micro-agent in the images consists of translation and rotation transformations. The agent is successfully tracked using the proposed visual tracking method. During the tracking, images are acquired at 20 fps. Fig. 6 shows results of visual tracking.

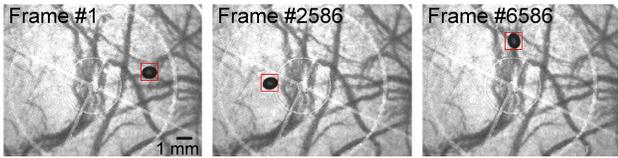


Fig. 6: Tracking results for a 1 mm elliptical shaped micro-agent in microscope images. Please refer to the accompanying video.

Tracking results for a 1 mm spherical micro-agent in microscope images are shown in Fig. 7. Similar to the previous experiment, the Petri dish is put on a printed vein pattern. Images are acquired at 25 fps. Instead of moving the agents using magnetic fields, the Petri dish is moved by hand to test the tracking robustness under large motion displacements. Also, this image sequence suffers from strong illumination variations. The center image shown in Fig. 7 suffers from strong light intensity and shadow. Further, the right image shown in Fig. 7 suffers from low intensity. The agent is successfully tracked under these conditions which confirms the robustness of the proposed visual tracking method.

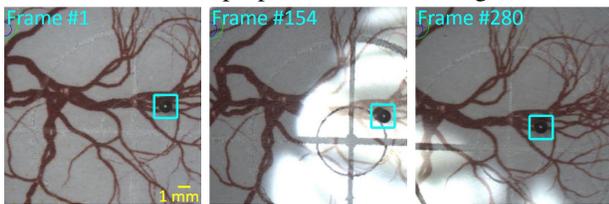


Fig. 7: Tracking results for a 1 mm spherical shaped micro-agent in microscope images under large motion displacements and significant illumination variations. Please refer to the accompanying video.

The proposed visual tracking method can also track multiple micro-sized agents. In our experiments, two spherically-shaped 100  $\mu\text{m}$  agents are moved independently using the controller described in [22]. The micro-agents are imaged using a microscope and tracked using both SSD and NCC based methods. During the tracking, images are acquired at 40 fps and the micro-agents are tracked in parallel. Tracking results are shown in the accompanying video.

2) *Multi-rate State Estimation Results:* The accuracy of multi-rate state estimation was evaluated using a high speed camera. We have implemented both Luenberger and Kalman state estimators offline. An elliptical shaped micro-agent with a length of 48 pixels was magnetically steered for 50 seconds. During the manipulation, frames were acquired at 200 fps and 9091 frames were captured in total. The accuracy of multi-rate state estimation is evaluated by tracking the particle at 200 fps which serves as ground truth. Then, multi-rate state estimation was performed by reducing the acquisition rate to 25, 50, 100 fps for  $N$  equals to 2, 4, and 8, respectively. Note that  $\mathbf{B}_d = \mathbf{0}$  and  $\mathbf{D}_d = \mathbf{0}$  holds for the system (25) and (26). Maximum absolute error (M.A.E) in  $x$ - and  $y$ - axes, mean and standard deviation of the error for different values of  $N$  and  $n_d$  as described in Section III-A are tabulated in Table I in pixels. An example output of multi-rate state estimation using Luenberger observer and Kalman filter in  $x$ - axes for  $N$  equal to 8 and  $n_d$  equal to 2 were depicted in Fig. 8. It can be observed from Table I, when  $n_d$  increases, the accuracy of the estimated state obtained using either a Kalman or a Luenberger state estimator and when  $N$  increases, the accuracy of the estimates reduces as expected.

TABLE I: Maximum absolute error (M.A.E), mean error, and standard deviation (std) are presented to compare performance between Luenberger and Kalman multi-rate state estimators. All units are given in pixels.

		$n_d = 0$		$n_d = 1$		$n_d = 2$	
		(Luen.)	(Kalm.)	(Luen.)	(Kalm.)	(Luen.)	(Kalm.)
$N = 2$	M.A.E -X	1.203	0.848	0.688	0.842	0.656	0.808
	M.A.E -Y	1.364	0.977	0.794	0.759	0.761	0.745
	Mean	0.324	0.106	0.230	0.071	0.214	0.065
	Std	0.197	0.094	0.107	0.074	0.010	0.073
$N = 4$	M.A.E -X	3.952	1.779	1.211	1.222	0.721	0.864
	M.A.E -Y	2.590	1.676	1.082	1.310	0.765	1.360
	Mean	0.946	0.333	0.334	0.166	0.236	0.110
	Std	0.746	0.301	0.195	0.164	0.113	0.125
$N = 8$	M.A.E -X	12.022	6.196	3.908	4.806	1.793	2.273
	M.A.E -Y	6.883	4.496	3.093	4.060	1.186	2.432
	Mean	3.237	1.231	0.959	0.819	0.403	0.355
	Std	2.441	1.109	0.733	0.793	0.305	0.394

3) *Collaborative Tracking Results:* A 100  $\mu\text{m}$  spherically shaped micro-agent is tracked with SSD and NCC based visual tracking methods in parallel. After each registration of template and current images, the outputs are merged separately by Luenberger (29) and Kalman (34), (35), (37), (38) estimators. Further, the merging is also applied to the case of multi-rate sampling. The results are shown in Fig. 9.

## VI. CONCLUSIONS

This study presents multi-rate Luenberger and Kalman state estimators for visual tracking of magnetic micro-agents. The main contribution is the intersample state estimation of micro-agents using slow medical imaging modalities. The

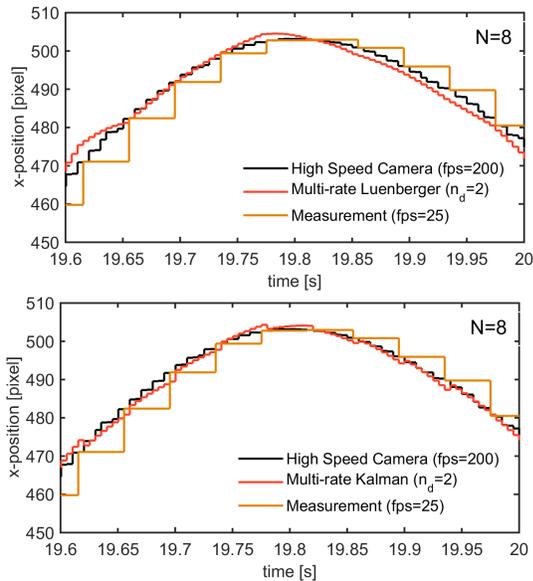


Fig. 8: Multi-rate sampling results using a Luenberger observer (top) and a Kalman filter (bottom) along the  $x$ -axis for  $N$  equals to 8 and  $n_d$  equal 2.

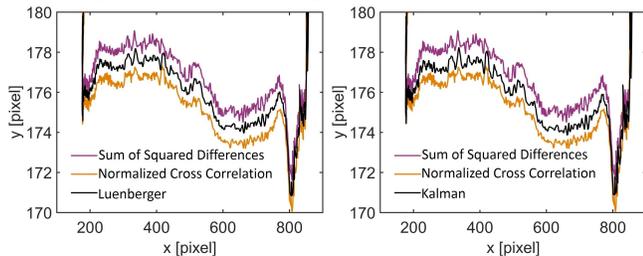


Fig. 9: Results of the merging process using Luenberger (left) and Kalman (right) state estimators.

micro-agents are tracked using SSD and NCC cost functions in medical images. Further, the outputs of SSD and NCC are merged using state estimators for robust tracking. During the experiments, magnetic micro-agents with different shapes are steered using coils and imaged separately with a 2D US machine and a microscope. The advantages of the proposed method are demonstrated by means of extensive experiments under challenging conditions such as strong illumination and intensity variations and large motion displacements. During the experiments, it was observed that Kalman filter was more accurate than Luenberger observer and the agents could be tracked more reliably using NCC based visual tracking compared to SSD based visual tracking. Experimental results demonstrate that the proposed method can accurately track micro-agents with different shapes in images obtained from slow medical imaging modalities while providing intersample estimates in real-time. We envision that this method could be readily utilized with clinical instrumentation to accelerate translation of the use of microrobots in realistic minimally invasive operations.

#### REFERENCES

- [1] B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott, "Microrobots for minimally invasive medicine," *Annual review of biomedical engineering*, vol. 12, pp. 55–85, 2010.
- [2] S. Scheggi, K. K. T. Chandrasekar, C. Yoon, B. Sawaryn, G. van de Steeg, D. H. Gracias, and S. Misra, "Magnetic motion control and planning of untethered soft grippers using ultrasound image feedback,"

- in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 6156–6161.
- [3] J. D. Keuning, J. de Vries, L. Abelmann, and S. Misra, "Image-based magnetic control of paramagnetic microparticles in water," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 421–426.
- [4] A. Hong, B. Zeydan, S. Charreyron, O. Ergeneman, S. Pan, M. F. Toy, A. J. Petruska, and B. J. Nelson, "Real-time holographic tracking and control of microrobots," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 143–148, Jan 2017.
- [5] S. Martel, O. Felfoul, J.-B. Mathieu, A. Chanu, S. Tamaz, M. Mohammadi, M. Mankiewicz, and N. Tabatabaei, "MRI-based medical nanorobotic platform for the control of magnetic nanoparticles and flagellated bacteria for target interventions in human capillaries," *The International journal of robotics research*, vol. 28, no. 9, pp. 1169–1182, 2009.
- [6] E. Diller, J. Giltinan, and M. Sitti, "Independent control of multiple magnetic microrobots in three dimensions," *The International Journal of Robotics Research*, vol. 32, no. 5, pp. 614–631, 2013.
- [7] S. Vaezy and V. Zderic, "Image-guided therapy systems," in *Medical Imaging*. Artech House, 2009, ch. 2.
- [8] Y. Wang, B. M. Nguyen, H. Fujimoto, and Y. Hori, "Multirate estimation and control of body slip angle for electric vehicles based on onboard vision system," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 2, pp. 1133–1143, Feb 2014.
- [9] M. Zheng, L. Sun, and M. Tomizuka, "Multi-rate observer based sliding mode control with frequency shaping for vibration suppression beyond nyquist frequency," *IFAC-PapersOnLine*, vol. 49, no. 21, pp. 13 – 18, 2016, 7th IFAC Symposium on Mechatronics Systems MECHATRONICS 2016.
- [10] M. Nemani, T.-C. Tsao, and S. Hutchinson, "Multi-rate analysis and design of visual feedback digital servo-control system," *Journal of dynamic systems, measurement, and control*, vol. 116, no. 1, pp. 45–55, 1994.
- [11] J.-Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 5, no. 1-10, p. 4, 2001.
- [12] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, Feb 2004.
- [13] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sept 2004, pp. 943–948.
- [14] G. R. Richa, R. Sznitman, and G. Hager, "Robust similarity measures for gradient-based direct visual tracking," *Technical report*, 2012.
- [15] M. Hwangbo, J. S. Kim, and T. Kanade, "Inertial-aided klt feature tracking for a moving camera," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 1909–1916.
- [16] L. Matthews, T. Ishikawa, and S. Baker, "The template update problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 810–815, June 2004.
- [17] R. A. DeCarlo, "Some techniques for computing the matrix exponential and its integral," in *Linear Systems: A State Variable Approach with Numerical Implementation*. Prentice-Hall, 1989, ch. 14.
- [18] H. Sira-Ramírez and V. F. Batlle, "Robust  $\Sigma$ - $\Delta$  modulation-based sliding mode observers for linear systems subject to time polynomial inputs," *International Journal of Systems Science*, vol. 42, no. 4, pp. 621–631, 2011.
- [19] R. Gahleitner and M. Schagerl, "A state observer for using a slow camera as a sensor for fast control applications," vol. 8661, 2013, pp. 8661 – 8661 – 10.
- [20] Z. P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857 – 869, 2001.
- [21] L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu, "3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.
- [22] A. Denasi and S. Misra, "Independent and leader–follower control for two magnetic micro-agents," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 218–225, 2018.
- [23] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 2756–2759.