

# Real Time Motion Analysis in 4D Medical Imaging Using Conditional Density Propagation

Johannes Lotz<sup>a</sup>, Bernd Fischer<sup>a</sup>, Janine Olesch<sup>a</sup> and Matthias Günther<sup>b</sup>

<sup>a</sup>Fraunhofer MEVIS Projectgroup Image Registration, Lübeck, Germany

<sup>b</sup>Fraunhofer MEVIS, Bremen, Germany

## ABSTRACT

Motion, like tumor movement due to respiration, constitutes a major problem in radiotherapy and/or diagnostics. A common idea to compensate for the motion in 4D imaging, is to invoke a registration strategy, which aligns the images over time. This approach is especially challenging if real time processing of the data and robustness with respect to noise and acquisition errors is required.

To this end, we present a novel method which is based only on selected image features and uses a probabilistic approach to compute the wanted transformations of the 3D images. Moreover, we restrict the search space to rotation, translation and scaling.

In an initial phase, landmarks in the first image of the series have to be identified, which are in the course of the scheme automatically transferred to the next image. To find the associated transformation parameters, a probabilistic approach, based on factored sampling, is invoked. We start from a state set containing a fixed number of different candidate parameters whose probabilities are approximated based on the image information at the landmark positions. Subsequent time frames are analyzed by factored sampling from this state set and by superimposing a stochastic diffusion term on the parameters.

The algorithm is successfully applied to clinical 4D CT data. Landmarks have been placed manually to mark the tumor or a similar structure in the initial image whose position is then tracked over time. We achieve a processing rate of up to 12 image volumes per second. The accuracy of the tracking after five time steps is measured based on expert placed landmarks. We achieve a mean landmark error of less than 2 mm in each dimension in a region with radius of 25 mm around the target structure.

**Keywords:** motion correction, real time, medical imaging, 4D CT, factored sampling

## 1. INTRODUCTION

Motion analysis in medical images that are acquired over time such as 4D CT or 4D ultrasound has gained importance for instance in radiotherapy<sup>1,2</sup> or for motion correction in combination with other modalities.<sup>3</sup> The large amount of data produced in 4D imaging imposes an efficient representation and also processing of the image data. In many cases, especially if real time processing of the data is necessary, the application of classical high accuracy methods, like for instance image registration, is very difficult due to their high computational cost.

Additionally, short acquisition times of the individual time frames which are necessary to obtain a high time resolution, weaken the quality of the acquired images and can cause artifacts and noise. Hence, an ideal motion analysis method also needs to be robust with regard to clutter and measurement errors.

The method proposed here is based on previous work<sup>4</sup> where conditional density propagation based on active contour features has been applied to 2D ultrasound data. We not only extend this scheme to 3D volume sequence data but in addition make use of pure landmark information. As an immediate consequence the underlying “condensation algorithm”<sup>5</sup> has to be implemented on an extended state space.

To evaluate the proposed method, it is applied to clinical lung tumor data and evaluated with respect to expert chosen landmarks.

## 2. APPROACH

### 2.1 Transformation Model

The overall goal of the algorithm is to estimate the motion of selected image features or of the entire image. We use an Eulerian framework and describe the image deformation as a parametric deformation of the underlying grid. To obtain voxel values at transformed grid points a nearest neighbor interpolator has been chosen because of its low computational cost. In order to compensate for respiratory volume change, we augment the common rigid transformation model with its six degrees of freedom (rotation around three axes and translation in x, y and z direction) by scaling in each coordinate direction. In a three-dimensional domain  $\Omega \subset \mathbb{R}^3$ , the transformation of a spatial coordinate  $u \in \Omega$  can be written as

$$u_{\text{trans}} = RAu + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \quad \text{where } A = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \text{ is a scaling matrix and}$$

$$R = R_x R_y R_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi_x) & -\sin(\varphi_x) \\ 0 & \sin(\varphi_x) & \cos(\varphi_x) \end{bmatrix} \begin{bmatrix} \cos(\varphi_y) & 0 & \sin(\varphi_y) \\ 0 & 1 & 0 \\ -\sin(\varphi_y) & 0 & \cos(\varphi_y) \end{bmatrix} \begin{bmatrix} \cos(\varphi_z) & -\sin(\varphi_z) & 0 \\ \sin(\varphi_z) & \cos(\varphi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is composed of three rotation matrices that correspond to the three rotation axes.

Compared to using a fully affine transformation with twelve degrees of freedom, the above transformation model has some advantages. The decomposition in elementary rotations, translations and scaling can be used to easily constrain the transformation to a physically meaningful range of parameters. Furthermore, the lower number of only nine free parameters reduces the computational cost of the parameter estimation.

The landmarks are defined such that the origin of the domain  $\Omega$  is located at the landmark's centroid and transformations are calculated with respect to this location.

### 2.2 Parameter Estimation

#### 2.2.1 Motion Modeling with Uncertainty

Due to the presence of noise in the measurement, our method is based on a probabilistic approach that includes a certain amount of uncertainty. In the following section, we closely follow the theory developed by Isard and Blake.<sup>5,6</sup> The authors suggest to model the motion process over time as a Markov chain. At a given time  $t$ , the random variable  $Z_t$  represents the information retrieved from the three dimensional image array  $\mathbf{I}$  and  $X_t$  represents the image transformations. A realization of these random variables corresponds to a measurement of an image volume, represented by a data vector  $\mathbf{z}_t$  with probability  $p(Z_t = \mathbf{z}_t | X_t = \mathbf{x}_t) =: p(\mathbf{z}_t | \mathbf{x}_t)$  given transformation parameters  $\mathbf{x}_t$ .

In order to approximate the actual motion in a measurement with respect to the previous timestep, the conditional density of the state  $\mathbf{x}_t$  is of interest. With the assumption of a Markov process in mind, the propagation over time can be expressed with Bayes rule as

$$p(\mathbf{x}_t | \mathbf{z}_t, \mathbf{z}_{t-1}) = k_t p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{t-1}) \quad (1)$$

with a normalization constant  $k_t$ .<sup>5</sup> The first term,  $p(\mathbf{z}_t | \mathbf{x}_t)$ , represents the measurement acquired from the image at the current time step. The second term can be interpreted as a prediction of the parameter state  $\mathbf{x}_t$  based on the density  $p(\mathbf{x}_{t-1} | \mathbf{z}_{t-1})$  of the previous time step. Assuming that all information from the observed image  $\mathbf{z}_{t-1}$  about the transformation parameters  $\mathbf{x}_t$  is represented in  $\mathbf{x}_{t-1}$ ,  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  can be used as an approximation of  $p(\mathbf{x}_t | \mathbf{z}_{t-1})$ . Here we model the change of  $X_t$  over time by means of a diffusion process<sup>5</sup> which in turn implies that  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  is of Gaussian nature.

### 2.2.2 Factored Sampling and CONDENSATION

The observation densities are not known analytically. However, we make use of the following approximation scheme. Given an observed image  $\mathbf{z}_t$ , factored sampling<sup>7</sup> is used to approximate the densities  $p(\mathbf{x}_t|\mathbf{z}_t)$  by sets of states  $\mathbf{s}_t = \{\mathbf{s}_t^i, i = 1, \dots, N\}$  sampled from  $X_t$  and weights  $\pi_t = \{\pi_t^i, i = 1, \dots, N\}$  that estimate the sample’s densities. These weights are calculated based on the image information by

$$\pi_t^i = \frac{p(\mathbf{z}_t|X_t = \mathbf{s}_t^i)}{\sum_{k=1}^N p(\mathbf{z}_t|X_t = \mathbf{s}_t^k)}. \quad (2)$$

The observation probability of transformed bright and dark landmarks  $u_b^k \in \Omega, k = 1, \dots, M_b$  and  $u_d^k \in \Omega, k = 1, \dots, M_d$  for a sample  $\mathbf{s}_t^i$

$$p(Z_t = \mathbf{I}|X_t = \mathbf{s}_t^i) = \left( \frac{1}{M_b} \sum_{k=1}^{M_b} \mathbf{I}(u_b^k) - \frac{1}{M_d} \sum_{k=1}^{M_d} \mathbf{I}(u_d^k) \right)^2 \quad (3)$$

is estimated by the difference of gray values in the observed image volume at these landmarks.<sup>4</sup>

Given the tuple  $(\mathbf{s}_{t-1}, \pi_{t-1})$  at time  $t - 1$ , the propagation step outlined in Equation (1) is implemented as follows.

1. A first approximation of  $\mathbf{s}_t$  is obtained by factored sampling from  $\mathbf{s}_{t-1}$ . We choose  $N$  times with repetition from  $\mathbf{s}_{t-1}$  such that the element  $\mathbf{s}_{t-1}^k$  is chosen with probability  $\pi_{t-1}^k$ . Samples with high probabilities are likely to be chosen more than once.
2. From the preliminary sample set  $\mathbf{s}_t$ , the distribution  $p(\mathbf{x}_t|\mathbf{z}_{t-1})$  can be obtained by random sampling from  $p(\mathbf{x}_t|X_{t-1} = \mathbf{s}_{t-1})$  which, by assuming a diffusion process over time, is Gaussian. With an additional drift towards the expectation value  $\bar{\mathbf{s}}_{t-1} = \sum_{i=1}^N \pi_{t-1}^i \mathbf{s}_{t-1}^i$  the diffusion update step is written as

$$\mathbf{s}_t^i = \alpha \mathbf{s}_t^i + (1 - \alpha) \bar{\mathbf{s}}_{t-1} + \sigma \mathbf{w}, \mathbf{w} \sim G(0, 1)$$

where the parameter  $\alpha$  is used to control the stabilizing effect of the drift term and  $\sigma$  defines the range in which the parameters are expected to change between two time steps.

3. To include the information from the acquired image at time  $t$ , the observation step from Equations (2) and (3) is applied to update the probability estimates  $\pi_t$  based on the new samples  $\mathbf{s}_t$ . The resulting tuple  $(\mathbf{s}_t, \pi_t)$  is used as initial value for the next iteration. The expectation value  $\bar{\mathbf{s}}_t$  is the estimated parameter vector of the current iteration.

The algorithm has been integrated into the image processing framework MeVisLab as a Python module. Additionally, the code has been rewritten in C++ to further reduce the computation time.

To evaluate the accuracy and timing, the algorithm has been applied to six test cases, each consisting of a time series of 3D CT lung images. These cases have been obtained from the DIR-lab datasets<sup>8</sup> which are publicly available and for which ground truth deformation data is available. From each time series we use the first six images, representing one full expiration. For the first and the sixth time step, 300 expert set landmarks are available and have been used as a ground truth to evaluate the calculated transformation. In the first time step, an exposed structure was marked by manually chosen landmarks which are divided into two groups. One group of landmarks is placed on bright tissue of the structure, the other group marks the dark surrounding area. This initial state is shown in an exemplary dataset in the top row of Figure 1. The tracking algorithm has been applied with the following parameters:

$N$	$\alpha$	$\sigma_{rotation}$	$\sigma_{translation}$	$\sigma_{scaling}$
500	0.9	5°	[3, 3, 7] mm	0.1

### 3. RESULTS

#### 3.1 Computation Time

We run the algorithm single threaded in C++ on an Intel Xeon E5645 multicore system with 2.40GHz. In all test cases, more than six image volumes per second can be processed. Table 1 shows the detailed timings of the individual test cases.

Timings per test case.

dataset	# bright landmarks	# dark landmarks	timing
1	22	10	7 FPS
2	11	7	8 FPS
3	9	6	9 FPS
4	9	7	12 FPS
5	16	15	6 FPS
6	26	19	5 FPS

Table 1. Timing results for six test cases. Sample size  $N = 500$ . Since the computation time strongly depends on the number of tracked landmarks, these numbers are given next to the computational speed expressed in processed image volumes per second. The sizes of the image volume range from  $512 \times 512 \times 128$  (case 6) to  $256 \times 256 \times 94$  (case 1). The number of landmarks needed for tracking depends mostly on the size of the tracked structure.

The results in Table 1 clearly show that the computation time mostly depends on the number of landmarks that have been chosen for tracking. Since the number of image access operations does not depend on the size of the image volume, this number is less important in the context of computational time.

#### 3.2 Accuracy

The accuracy of the algorithm is measured based on expert set landmarks which are contained in the DIR-lab dataset. In the image volume at time zero, we chose a region with a radius of 25 mm around the center of the tracked structure. The landmarks in this region are transformed using the parameters calculated by our algorithm at the sixth time step (maximum expiration) and compared to the corresponding expert landmarks. Figure 1 shows a CT image with transformed landmarks at three of the first six time points. The landmark errors are shown in Table 2.

Landmark error towards ground truth expert landmarks.

dataset	error $e_x$	error $e_y$	error $e_z$	stdev $s_x$	stdev $s_y$	stdev $s_z$	n
1	1.0 mm	1.2 mm	1.4 mm (0.6 vx)	0.3 mm	0.2 mm	1.1 mm	3
2	1.3 mm	0.7 mm	0.3 mm (0.1 vx)				1
3	1.3 mm	1.1 mm	2.0 mm (0.8 vx)	0.8 mm	0.6 mm	1.4 mm	5
4	3.3 mm	1.0 mm	3.2 mm (1.3 vx)	0.5 mm	0.4 mm	0.3 mm	6
5	1.5 mm	1.1 mm	1.3 mm (0.5 vx)	0.8 mm	0.9 mm	1.3 mm	9
6	1.2 mm	1.7 mm	2.1 mm (0.8 vx)	0.8 mm	1.1 mm	1.5 mm	9
mean	1.6 mm	1.1 mm	1.7 mm (0.7 vx)				

Table 2. Accuracy results for six test cases. Sample size  $N = 500$ . Columns  $e_x$ ,  $e_y$  and  $e_z$  show the mean of the absolute errors in each dimension calculated by transforming ground truth expert landmarks. The standard deviation of the errors is shown in columns  $s_x$ ,  $s_y$  and  $s_z$  respectively. The number of ground truth landmarks in the region with radius 25 mm around the center of the analyzed structure is shown in column n.

### 4. DISCUSSION

We present a novel registration scheme which is capable to facilitate a proper 3D motion tracking in real time. The proposed scheme turns out to be considerably faster as compared to conventional intensity based registration schemes, while the landmark accuracy deteriorates only slightly (compare<sup>8,9</sup>).

The algorithm is especially suited for time-critical or real-time applications. Opposed to derivative based methods, where the time to convergence depends heavily on the input data, the computation time of the presented tracking algorithm mostly depends on the number of chosen landmarks and the sample size  $N$ . Therefore, the computation time can be estimated before the computation. Furthermore, a trade-off between computational speed and accuracy is possible by reducing or increasing the sample size  $N$ .

The dependence of accuracy and computation speed on the number of samples is shown in Figure 2. The variance in the results decreases with increasing sample size. Due to the probabilistic nature of the algorithm, multiple executions, even on the same data, lead to different results. For each tested sample size, some of the runs result in a low error. However, for small sample sizes, the risk to obtain a false or inaccurate result increases dramatically due to the higher variance in the results.

It is worth noticing that the presented implementation has not yet been optimized for speed and only runs single threaded. We expect a further speedup from parallelization on CPU or GPU as it has been already demonstrated in the 2D case.<sup>4</sup>

Our results show that the presented scheme produces accurate results in a limited region around the center of the tracked structure. If more distant regions are included in the calculation of the error, the measured accuracy is reduced. This indicates that for more detailed motion analysis of larger structures, more complex and in consequence more time consuming methods are needed. Various approaches to compute a non-linear deformation of the whole lung have been presented<sup>9,10</sup> for the same data set.

The current implementation still requires a manual user intervention to select the structure to be tracked and to choose appropriate landmark positions. Much of the accuracy of the algorithm depends of the accurate placement of the tracking landmarks. The presented scheme would benefit from a combination with an automatic landmark selection method which we will investigate in future work.

## REFERENCES

- [1] Murphy, M. J., “Tracking moving organs in real time,” *Seminars in Radiation Oncology* **14**, 91–100 (Jan. 2004).
- [2] Sharp, G. C., Jiang, S. B., Shimizu, S., and Shirato, H., “Prediction of respiratory tumour motion for real-time image-guided radiotherapy,” *Physics in Medicine and Biology* **49**, 425–440 (Feb. 2004).
- [3] Günther, M. and Feinberg, D. A., “Ultrasound-guided MRI: preliminary results using a motion phantom.,” *Magnetic Resonance in Medicine* **52**, 27–32 (July 2004).
- [4] Zhang, X., Günther, M., and Bongers, A., “Real-time organ tracking in ultrasound imaging using active contours and conditional density propagation,” *Medical Imaging and Augmented Reality* , 286–294 (Sept. 2010).
- [5] Isard, M. and Blake, A., “Condensation — Conditional density propagation for visual tracking,” *International Journal of Computer Vision* **29**, 5–28 (Aug. 1998).
- [6] Blake, A. and Isard, M., [*Active contours*], Springer, London (Apr. 1998).
- [7] Grenander, U., Chow, Y., and Keenan, D. M., [*Hands: A pattern theoretic study of biological shapes*], Springer, New York (Jan. 1991).
- [8] Castillo, E., Castillo, R., Martinez, J., Shenoy, M., and Guerrero, T., “Four-dimensional deformable image registration using trajectory modeling,” *Physics in Medicine and Biology* **55**, 305–27 (Jan. 2010).
- [9] Castillo, E., Castillo, R., White, B., Rojo, J., and Guerrero, T., “Least median of squares filtering of locally optimal point matches for compressible flow image registration.,” *Physics in Medicine and Biology* **57**, 4827–33 (Aug. 2012).
- [10] Rühaak, J., Heldmann, S., Kipshagen, T., and Fischer, B., “Highly Accurate Fast Lung CT Registration,” *Proceedings of SPIE 8669, Medical Imaging 2013: Image Processing* (Mar. 2013).

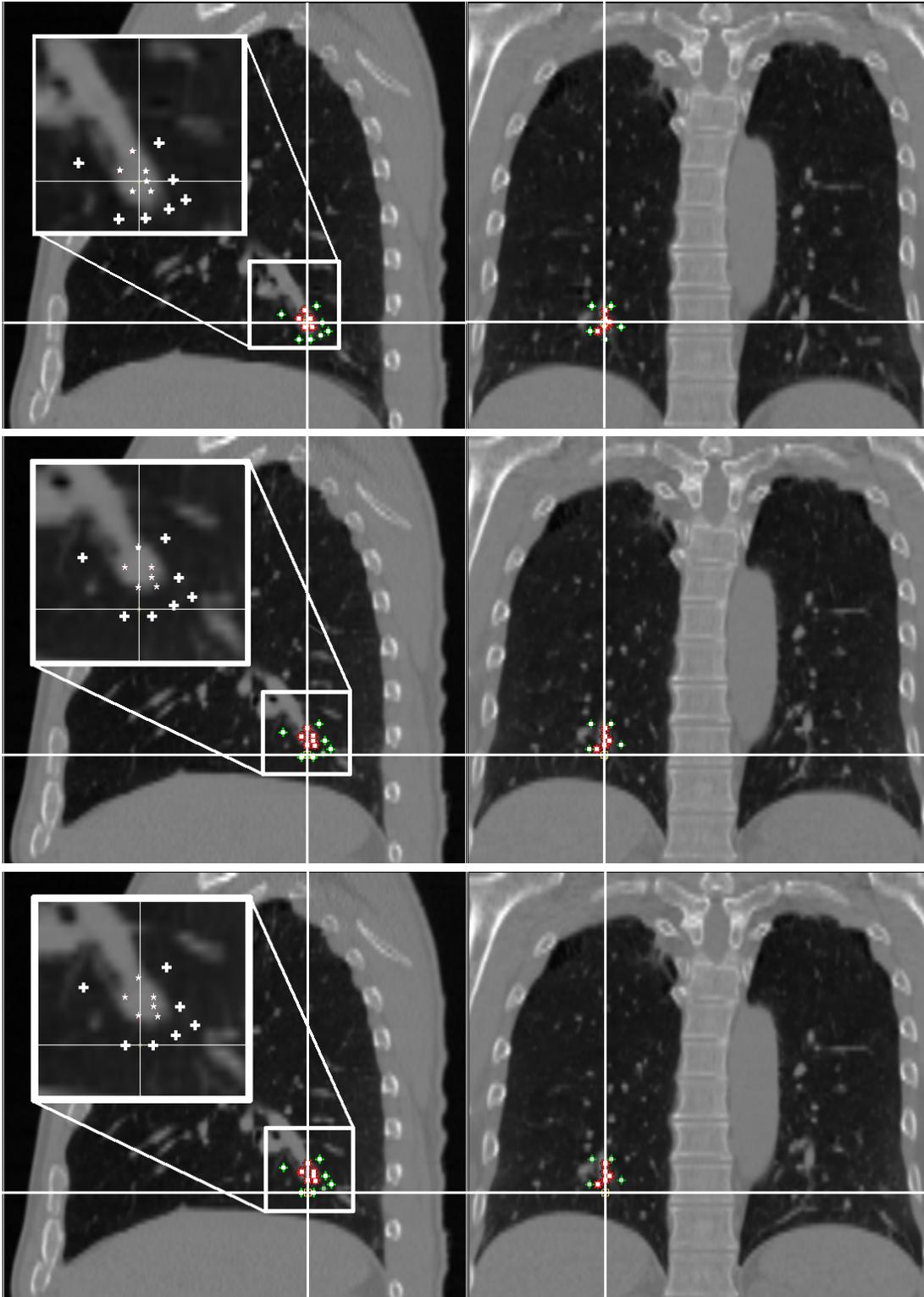


Figure 1. Case 2. 4D lung CT image with tracked structure and automatically following landmarks in an upwards motion. Crosshairs marks the initial position of the structure for comparison. From top to bottom: Slices through the image volumes at time points 1, 3 and 6 with landmarks transformed as a result of the presented tracking algorithm. Left: coronal, center: sagittal view.

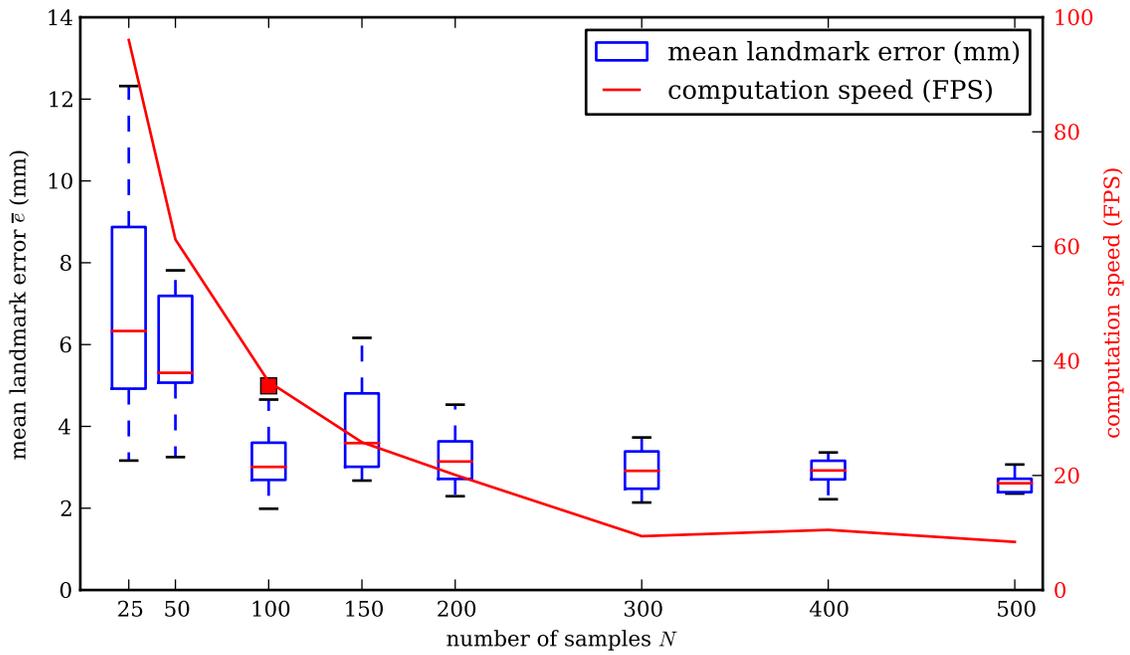


Figure 2. Sample size  $N$  plotted against mean landmark error (box-whisker plots, left scale) and computation speed (solid line, right scale) evaluated on test case 5 using 28 landmarks. The algorithm has been run ten times for each of the eight different choices of  $N$ . The box-whisker-plots show the distribution of the mean errors measured after each execution. The computation speed shown is an average taken from three subsequent executions of the algorithm for each  $N$ .