Subject-specific liver motion modeling in MRI

Document status and date:
Published: 07/04/2017

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Subject-specific liver motion modeling in MRI: a feasibility study

Yolanda H Noorda, Lambertus W Bartels, Max A Viergever and Josien P W Pluim

Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands

E-mail: `yolanda@isi.uu.nl`

7 September 2016

Abstract. A liver motion model based on registration of dynamic MRI data, as previously proposed by the authors, was extended with temporal prediction and respiratory signal data. The potential improvements of these extensions with respect to the original model were investigated. Additional evaluations were performed to investigate the limitations of the model regarding temporal prediction and extreme breathing motion.

Data were acquired of four volunteers, with breathing instructions and a respiratory belt. The model was built from these data using spatial prediction only and using temporal forward prediction of 300 ms to 1200 ms.

From temporal prediction of 0 ms to 1200 ms ahead, the Dice coefficient of liver overlap decreased with 0.85%, the median liver surface distance increased with 20.6% and the vessel misalignment increased with 20%. The mean vessel misalignment was 2.9 mm for the original method, 3.42 mm for spatial prediction with a respiratory signal and 4.01 mm for prediction of 1200 ms ahead with a respiratory signal.

Although the extension of the model to temporal prediction yields a decreased prediction accuracy, the results are still acceptable. The use of the breathing signal as input to the model is feasible. Sudden changes in the breathing pattern can yield large errors. However, these errors only persist during a short time interval, after which they can be corrected automatically. Therefore, this model could be useful in a clinical setting.

Keywords: MR-HIFU, liver motion model, registration, dynamic data, temporal prediction, Kalman filtering

1. Introduction

Organ motion modeling can be valuable for image-guided non-invasive treatments. For the liver in particular, many models exist. Examples are population-based models [1, 2] and subject-specific models [3, 4]. Possible applications for liver motion models are radiotherapy procedures [5] and Magnetic Resonance-guided High Intensity Focused Ultrasound (MR-HIFU) treatment of liver lesions. In both methods, compensation for

liver motion is required. This can be done by performing gated therapy, in which the treatment is only performed when the target is at a specific position. For MR-HIFU this approach is less efficient, due to the high vascularization of the liver. The high perfusion that results from this causes relatively fast cooling of the target. Therefore, a continuous treatment approach would be beneficial, in which the tumor position can be followed over time. We previously introduced a subject-specific motion model that was based on registration of dynamic MR data in order to create a look-up table describing the deformation of the liver in several states in an average breathing cycle [6]. This model performs spatial prediction of the average liver motion, by which we mean that there is no forward prediction in time, but the deformation at a specific time point is predicted by the model. This is sufficient for use during treatment planning, to investigate the feasibility of a procedure and to determine an optimal treatment strategy. For such a model to also be useful during the actual treatment, temporal prediction of the liver motion is required, in which the state of the liver is predicted for future time points, for which the deformation is then predicted by the model. This information can be used during treatment to adjust the geometry according to the new target location. In this work, both the spatial and temporal prediction results are evaluated, also for more realistic situations in which breathing can be irregular.

The building of the model is an extensive process including many registrations, which can however be performed completely offline. During treatment, to link the actual motion to the model, input would be needed on the breathing pattern and the current phase in the breathing cycle, which can be obtained with a respiratory belt or with a navigator echo applied during MR acquisition. Based on this information, the liver deformation corresponding to the current state could then be obtained from the model.

In this work, we investigate the spatiotemporal performance of our model, for four volunteers. This study was performed to yield a better insight into the potential of our model to be applied in clinical practice as well as into further extensions that might be needed. We acquired a breathing signal using a respiratory belt during the MR data acquisition, which yields the information necessary for temporal motion prediction. In addition to regular breathing, volunteers were also instructed to lower or increase their breathing frequency and to perform deep inhalations or shallow breathing, to explore the limitations of our motion prediction method regarding variations in breathing pattern. Experiments were set up to determine the influence of the number of time prediction steps on the accuracy of the model.

Materials & Methods

Data

MR data were acquired of four healthy volunteers. In accordance with local regulations in the hospital, the volunteers were screened for contraindications and asked to sign an

Table 1: MR data acquired on volunteers.

Scan parameters	Breathhold volume scan	Multi-slice dynamic scan	Single-slice dynamic scan
Acquisition	3D	Multi-slice 2D $+ t$	Single-slice 2D $+ t$
Orientation	Transversal	Sagittal	Sagittal
Technique	T1-TFE	T1-TFE	T1-TFE
TE (ms)	1.8	1.8	1.8
TR (ms)	3.9	3.9	3.9
Flip angle(°)	30	30	30
Voxel size in-plane (mm^2)	2×2	2×2	2×2
Slice thickness (mm)	4	8	8
Slice gap (mm)	-2	15	N/A
Acquisition matrix	256×256	256×256	256×256
Number of slices	100	6	1
Number of dynamics	N/A	200	150
Half-scan in-plane factor	0.7	0.7	0.7
Half-scan through-plane factor	0.85	0.85	0.85
Fat suppression	SPAIR	SPAIR	SPAIR
Total scan duration (s)	22.5	369	46.2

informed consent form. All MRI scans were acquired on a 1.5-T scanner (Achieva, Philips Healthcare Best, The Netherlands). The volunteers were scanned in prone position. For each volunteer, a 3D transverse scan was acquired during a breathhold at inspiration. In addition, a series of 200 2D dynamically acquired sagittal multi-slice scans with 6 slices was recorded. The six slices were placed at equidistant locations in the liver, with a gap of 15 mm in between them, as shown in Figures 1 and 2. The dynamics were acquired interleaved over these six locations. During this scan, the volunteers were given breathing instructions to obtain the following breathing modes: fast breathing (0.60-0.80 Hz), slow breathing (0.10-0.25 Hz), sudden breathhold and regular breathing (0.20-0.25 Hz). The volunteers were instructed to breathe normally for 10 s, then increase their breathing frequency for the next 10 s, breathe slowly for 10 s, perform a sudden breathhold of 5 s and then breathe normally for the last 10 s. Finally, a sagittal single 2D dynamic slice sequence was acquired, which was identical to the scan acquired with the dynamic protocol mentioned above with the same orientation. This series consisted of 150 dynamics and was acquired at a different location in the liver, for validation of the model. A breathing signal was recorded during all scans using a respiratory belt. In Table 1, an overview of the scan parameters for these three acquisitions is presented.

Methods

Building of the model   For all subjects, a liver motion model was built based on the 3D volume data and the multi-slice dynamic scan. This resulted in a look-up table typically
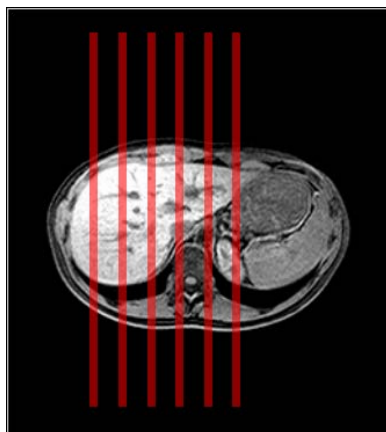
Figure 1: The geometry of the dynamic acquisition overlaid on the volume scan.
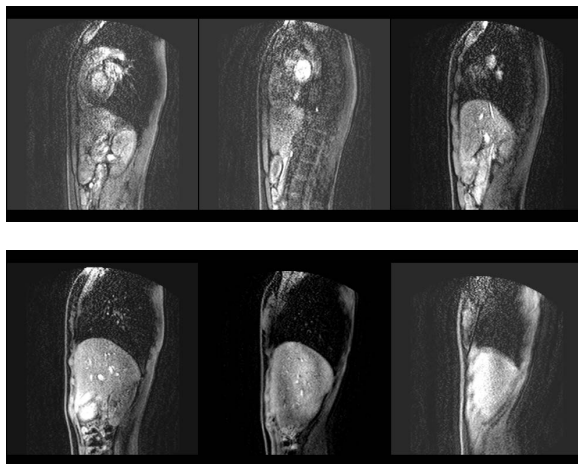

Figure 2: The six dynamic slices.

consisting of 20 possible liver states: 10 during inspiration and 10 during expiration. These liver states were represented by full 3D transformations, consisting of a rigid motion component and a deformation. It was assumed that the normalized feet-head (FH) translation of the liver, together with its direction, could be used to distinguish each liver state in the model. Based on this value, for every new liver image, the most similar state can be obtained from the model. Its corresponding 3D deformation field can then be used to deform the image according to this state. For details on the model construction, we refer to our earlier work [6].

In this work, the model was built in different configurations. First of all, the exact same method was used as described in our previous work [6], where a simulated navigator signal, since acquiring an actual navigator in the same imaging sequence that was used for dynamic imaging could not be realized, was used to measure the translation of the liver in feet-head (FH) direction and spatial motion prediction was performed (configuration N). In addition, the model was built using a recorded respiratory signal instead of a simulated navigator signal. In this configuration, not only spatial prediction

was performed, but temporal prediction was also performed for 1 to 4 timesteps ahead, corresponding to 300 to 1200 ms (configuration R). Finally, the model was also built in selective configuration, meaning that dynamics with deviating frequency or amplitude were not contained in the model (configuration RS). Samples were removed if the respiratory frequency (which was determined by a Fourier transform spectral analysis within a window of [-5s, +5s]) or the amplitude at that time point was more than one standard deviation above or below the average frequency or amplitude, respectively. As a result, on average 30% of the samples were removed (keeping 65.7%, 71.2%, 80.3% and 60.0% of the samples for subjects 1 to 4, respectively), which represented extremely deep inhalations or extremely fast or slow breathing. Samples with smaller irregularities were still contained in the dataset. This selective configuration was also combined with temporal prediction dt of 0 to 1200 ms ahead. All configurations are listed below.

- N0: Spatial prediction using a simulated navigator (as was done in [6]): $dt = 0$ ms.
- R0: Spatial prediction using the breathing signal: $dt = 0$ ms.
- R300: Temporal prediction using the breathing signal: $dt = 300$ ms (1 time step).
- R600: Temporal prediction using the breathing signal: $dt = 600$ ms (2 time steps).
- R900: Temporal prediction using the breathing signal: $dt = 900$ ms (3 time steps).
- R1200: Temporal prediction using the breathing signal: $dt = 1200$ ms (4 time steps).
- RS0: Spatial prediction using the breathing signal, model built in selective configuration: $dt = 0$ ms.
- RS300: Temporal prediction using the breathing signal, model built in selective configuration: $dt = 300$ ms (1 time step).
- RS600: Temporal prediction using the breathing signal, model built in selective configuration: $dt = 600$ ms (2 time steps).
- RS900: Temporal prediction using the breathing signal, model built in selective configuration: $dt = 900$ ms (3 time steps).
- RS1200: Temporal prediction using the breathing signal, model built in selective configuration: $dt = 1200$ ms (4 time steps).

The look-up table that was created by the model was used for motion prediction on the sagittal single slice dynamic series that was acquired for evaluation. This sequence was acquired at a location in the middle of the liver, which was not contained in the model. The input to the model was the normalized feet-head (FH) translation of the liver, together with its direction. Based on this value, the current state of the liver could be linked to the most similar state contained in the model.

Comparing the results of configuration N0 with R0 reflects the additional error that is introduced in the practical situation, which relies on the breathing signal instead of a simulated navigator signal. Comparing the selective configuration (RS) with the standard configuration (R) yields insight into the influence of differences in the breathing

pattern during the building of the model. Finally, the influence of the prediction time is analyzed by comparing the results of the configurations with different prediction times.

Using the model with a navigator signal (N-configuration) In the N-configuration, the FH-translation values were obtained by simulating a navigator echo, by sampling a 1D column in each dynamic and tracking the diaphragm in these images. The direction of the translation could be obtained from the gradient of this simulated signal. The normalized FH-translation value was calculated for each dynamic in the evaluation sequence. This value was then used to determine the corresponding liver state of these dynamics. Each normalized FH-translation value was compared with the average FH-translation value of each liver state and the state with the closest average value was assigned to the corresponding dynamic.

Using the model with a respiratory signal (R-configuration) To evaluate the performance of the model based on the respiratory signal, a relationship needed to be established between this signal and the FH-translation of the liver, to obtain phase coherence between the two signals. A breathing signal was obtained with an external respiratory sensor, which was placed posteriorly (since the subjects were scanned in prone position), approximately at the average diaphragm position. The position of this sensor was logged with a frequency of 500 Hz. Logging of the breathing signal starts before the start of the actual scan, which yields data for calibration of the signal to the model. Since the dynamic scan acquisition time was 300 ms, the original respiratory signal was downsampled to match this time step. Next, the average phase shift between the two signals computed using the data points acquired before the start of the scan. This shift was then applied to the evaluation respiratory signal. From this relationship, normalized FH-translation values were obtained. Inspiration states were again separated from expiration states using the gradient of the translation. The coupling with the model was then achieved in the same manner as with the simulated navigator signal.

Temporal motion prediction In addition to spatial prediction, temporal prediction of the liver motion was performed. The temporal resolution of the dynamic sequence was 300 ms. The limitations of the prediction were investigated by varying the prediction time. Results were generated for time steps of 0 ms (spatial prediction only), 300 ms, 600 ms, 900 ms and 1200 ms. This temporal prediction was performed using an extended Kalman filter, which has been shown to yield robust predictions for non-linear motion [7, 8, 9, 10, 11, 12]. The Kalman filter is a recursive prediction filter, that assumes the state of a process at a certain time point can be predicted based on its previous states. The extended Kalman filter is the non-linear variant, which locally uses a linear approximation of the motion. In this case, it was assumed the liver motion could be approximated by a driven harmonic oscillator, based on the assumption that its periodic motion is caused by the periodic pattern of the breathing [13, 14], which in turn induces

motion of the diaphragm. In this case, the following equation of motion applies:

$$\omega_0^2 x + \beta \dot{x} + \ddot{x} = f, \tag{1}$$

with x representing the position, ω_0 the angular frequency, β a friction coefficient and f the driving force. Following the approach in [9], the current state of the motion is described by the state vector $s_t = (x(t), v(t), f(t))$, describing the position, velocity and driving force at a certain time point. Using Equation 1, the time evolution can be written as:

$$\frac{\partial s}{\partial t} = As \tag{2}$$

Assuming $\dot{f} = 0$ in a short time frame, since nothing is known about the driving force, we obtain the matrix A:

$$\begin{pmatrix} 0 & 1 & 0 \\ -\omega_0^2 & -\beta & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Time evolution for a short time interval Δt is then given by the matrix exponential:

$$\phi(\Delta t) = e^{A\Delta t}. \tag{3}$$

Let $F := \phi(\Delta t)$ define the time evolution matrix for a certain time step. Then Equation 2 can be written in a recursive form as:

$$s_t = Fs_{t-1}. \tag{4}$$

In practice, an additive noise term w_t, describing the process noise, should be taken into account:

$$s_t = Fs_{t-1} + w_t. \tag{5}$$

This noise term is expected to be normally distributed with zero mean and covariance Q.

The Kalman filter can be used to predict the state vector s at a certain time point, based on the previous time point. The Kalman filter comprises two phases: the prediction phase and the update phase. In the prediction phase, an estimate of the state at the next time point is made, based on the model that was used as an input. Equation 4 describes the ideal case when no noise is present. This equation is used to predict the *a priori* state $s_{t|t-1}$ based on the previous state $s_{t-1|t-1}$. In addition, an *a priori* estimate of the covariance matrix P_t, reflecting the expected prediction error, is made. The prediction step thus contains the following two calculations:

$$\hat{s}_{t|t-1} = F\hat{s}_{t-1|t-1} \tag{6}$$
$$P_{t|t-1} = FP_{t-1|t-1}F^T + Q. \tag{7}$$

In our example, we defined Q as:

$$\begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

such that errors were attributed to the force more heavily than to the position and the velocity. This was done since this force should represent the causes of all unpredictable short-term additional variations in the motion pattern, such as surrounding organ motion and changes in the breathing pattern. These values were set after performing a few experiments on records that were not used for evaluation.

In the update step, the state vector is updated based on a new measurement, taking into account possible observation noise. The measured state z_t is described as:

$$z_t = H\hat{s}_t + r_t, \tag{8}$$

where H_t is the measurement matrix and r_t is the observation noise. This observation noise is also expected to be normally distributed with zero mean and covariance R. In this case, the measurement matrix H was given by $H = (1, 0, 0)$, since only position was measured. The observation error covariance R was set to $(0.2, 0, 0)$, reflecting the estimate that the error margin in measuring the normalized FH-translation might be approximately 20%. This should include both the error in the relationship between the breathing signal and the liver motion, as well as the error resulting from the registrations. Before each update, the residual y_t is calculated, which is the difference between the measured state and the state that would have been measured in case of an accurate prediction:

$$\tilde{y}_t = z_t - H\hat{s}_{t|t-1} \tag{9}$$

The residual covariance is given by:

$$S_t = HP_{t|t-1}H^T + R \tag{10}$$

The *a posteriori* state estimate given by the update step, is based on both the *a priori* estimate from the prediction step and the measured state in the update step. Based on the covariance matrices P and S, a weighting is found between both factors in order to yield a new state estimate. This balance, which reflects the confidence in the prediction versus the confidence in the measurements, is described by the optimal Kalman gain:

$$K_t = P_{t|t-1}H^T S_t^{-1} \tag{11}$$

Once the optimal Kalman gain is calculated, the update step gives the *a posteriori* state estimate and the *a posteriori* covariance estimate:

$$\hat{s}_{t|t} = \hat{s}_{t|t-1} + K_t\tilde{y}_t \tag{12}$$
$$P_{t|t} = (I - KH)P_{t|t-1}. \tag{13}$$

To predict more than one time step ahead, several consecutive prediction steps are performed for each time point, before an update step is performed.

Evaluation

The liver motion model yielded a look-up table, containing deformed liver volumes, representing liver motion and deformation throughout an average breathing cycle. From each of these deformed volumes, an interpolated slice was taken at the same location

as the dynamic evaluation sequence. This slice was then compared with the actual state of the liver, represented by the evaluation sequence. Quantitative evaluation was performed by computing the Dice coefficient between the actual and predicted liver mask, the median surface distance between the actual and predicted liver boundary and by comparing the predicted blood vessel center positions with the actual center positions. Since this process was time-consuming, this was done for 20% of the dynamics (30 dynamics) for every subject. The details are explained below.

Model performance in all configurations To evaluate the performance of the model quantitatively, semi-automatic segmentations of the liver were created directly on the dynamic evaluation slices, using an active contour algorithm [15], followed by manual correction for leaking. A liver segmentation was created in the same way on the interpolated slice taken from the original volume. This mask was then deformed according to the deformation fields yielded by the model, to generate the ground truth liver masks. The Dice coefficient between these masks was computed to compare their similarity.

The blood vessel misalignment was computed by comparing blood vessels in the actual dynamic image and the predicted image. Both images were binarized using a manually set threshold to yield a blood vessel segmentation. Only vessels that were visible in both images were compared. The number of vessels could differ per dynamic and consequently, evaluation was performed on a varying number of vessels (two to five). For each vessel, the center of mass was computed on both images, and the Euclidean distance between these two points was calculated to obtain the vessel misalignment in millimeters. Unfortunately, the shape of the vessel profile could be different between the two images, as a consequence of pulsatility and inflow effects, causing through-plane motion. Consequently, an additional error was introduced by comparing the centers of mass of the vessel profiles. This should be considered when evaluating the vessel misalignment.

The outcomes of the model in the different configurations were used to compare the performance of building the model based on a navigator signal (N0) or on the respiratory signal (R0). In addition, a comparison was made between spatial prediction based on the respiratory signal (R0/RS0) and prediction for 1 to 4 time steps ahead (R300, R600, R900, R1200/RS300, RS600, RS900, RS1200). Finally, a comparison was made between including all liver states that occurred during the building of the model (R-configuration) and by only selecting liver states that did not represent extreme situations (RS-configuration).

1.0.1. Separating internal errors in the model from Kalman prediction errors during its use The final prediction error of the liver motion model results from two independent errors: registration errors and errors in signal prediction by the Kalman filter. It is assumed that the recording of the breathing signal was sufficiently accurate, so that possible errors introduced by this process could be considered negligible. Registration

errors occur in the process of building the model. These errors result in incorrect FH-translation values, which cause errors in the assignment of a sample to a particular breathing state. Moreover, additional errors in the deformable registration step cause inaccuracies in the deformation fields of the cluster. Therefore, registration errors influence the accuracy of the look-up table.

Errors originating from the Kalman filter affect the model performance by introducing errors in the use of the look-up table. These errors cause an incorrect prediction of the respiratory phase and therefore the corresponding sample is linked to an incorrect FH-translation value. Consequently, the wrong cluster is assigned to this sample and the deformation applied at this time point is incorrect. Errors in signal prediction become significant when they are larger than half the distance between different model states. Typically, this distance was found to be approximately 20% of the normalized respiratory amplitude. The entire motion range is twice as large as the amplitude, therefore this corresponds to 10% of the normalized range.

To investigate the influence of the extended Kalman filter on the final model prediction error, a separate evaluation was performed on signal prediction. Three evaluations were performed on the Kalman filter:

- The Kalman prediction errors were computed for all evaluation sequences, for the time steps of 300, 600, 900 and 1200 ms. These errors reflect the additional temporal prediction error, on top of the internal spatial prediction error, that can be expected when using the model in a clinical setting. To measure the Kalman prediction error, the difference between the actual normalized respiratory position and the predicted normalized position was computed for all samples. Because both signals had been normalized, the absolute respiratory amplitude was measured for all subjects, so that the Kalman prediction error could be translated to a final prediction error in millimeters.

- The Kalman filter performance in extreme situations was analyzed, to investigate the maximal prediction error that can be expected in a practical situation. Therefore, the Kalman prediction error was computed on the outlier samples in the dataset that was used to build the model. These extreme situations were: a deep inhalation, a sudden change in breathing frequency, and a sudden stop. Again, the Kalman prediction errors for time steps of 300, 600, 900 and 1200 ms were computed.

- To investigate if a larger Kalman prediction error indeed leads to less accurate results in the end, the correlation coefficients between the Kalman filter error and the errors in the Dice coefficient, the median surface distance and the blood vessel positions were calculated. The corresponding P-values were computed to investigate the significance of the correlation between prediction errors in the Kalman filter and resulting errors in the final results. In addition, the influence on the final model prediction outcome was inspected, by separating the results of samples with a Kalman error below and above 0.2. A Kalman error of 0.2 was
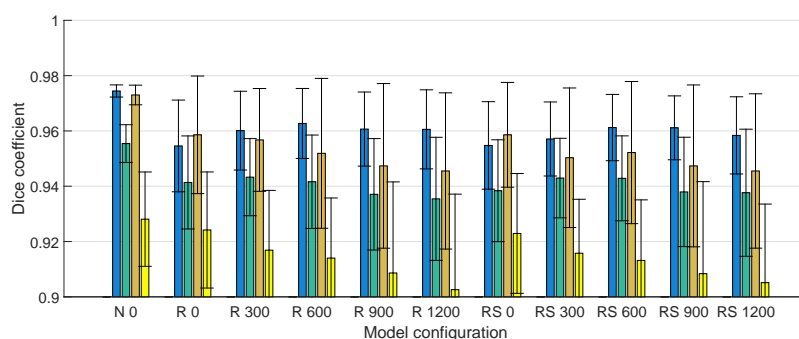
Figure 3: Dice coefficients of a comparison between the actual liver masks on 30 dynamic evaluation slices and the predicted liver masks. The different colors represent the four different subjects.
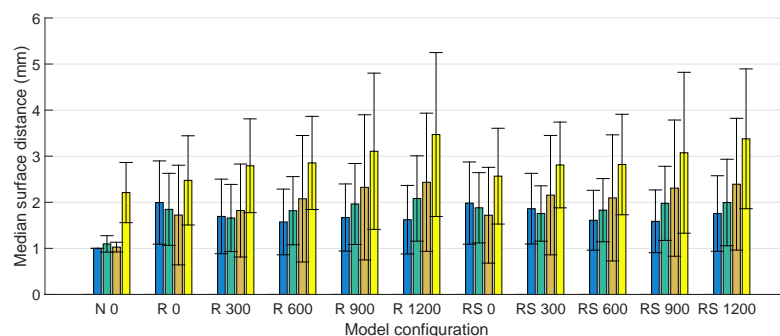

Figure 4: Median surface distance in mm between the actual liver boundaries on 30 dynamic evaluation slices and the predicted boundaries. The different colors represent the four different subjects.

assumed to correspond to the average distance in FH-position between different states in the model look-up table.

Results

Figure 3 shows the Dice coefficients obtained for all subjects, using the 11 different model configurations. For all subjects, method N0 yields the highest Dice scores, followed by method RS0.

In Figure 4, the median surface distance is shown for all subjects and model configurations. The best performing model configuration for the median surface distance is N0, followed by R0.

Figure 5 shows the mean vessel misalignment for all subjects and model configurations. The most eminent difference with the Dice score and the surface distance are the results for subject 1. Contrary to the positive results of these measures, the vessel misalignment is the poorest for this subject (4.18 - 4.70 mm). On average, the best performing model configuration is again N0, followed by RS0.
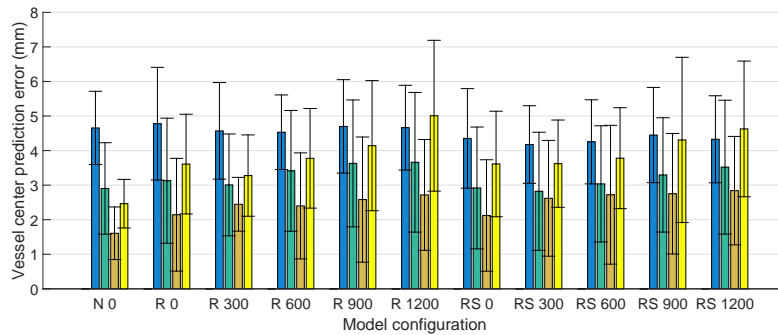
Figure 5: Vessel misalignment in mm between the actual vessel centers on the dynamic evaluation slices and their predicted positions. The different colors represent the four different subjects.

Table 2: Average and standard deviation of results for all subjects, computed for all model configurations.

Model configuration	Dice coefficient	Median surface distance (mm)	Vessel misalignment (mm)
N0:	0.958 ± 0.022	1.33 ± 0.59	2.91 ± 1.28
R0:	0.945 ± 0.016	2.01 ± 0.33	3.42 ± 1.09
R300:	0.944 ± 0.020	1.99 ± 0.54	3.33 ± 0.90
R600:	0.943 ± 0.021	2.08 ± 0.56	3.53 ± 0.89
R900:	0.938 ± 0.022	2.27 ± 0.62	3.76 ± 0.91
R1200:	0.936 ± 0.025	2.40 ± 0.79	4.01 ± 1.04
RS0:	0.944 ± 0.016	2.04 ± 0.37	3.25 ± 0.95
RS300:	0.942 ± 0.018	2.15 ± 0.47	3.31 ± 0.72
RS600:	0.942 ± 0.021	2.09 ± 0.53	3.45 ± 0.70
RS900:	0.939 ± 0.022	2.24 ± 0.63	3.70 ± 0.82
RS1200:	0.937 ± 0.023	2.38 ± 0.71	3.83 ± 0.81

Table 2 shows the performance for all model configurations. For all three measures, spatial prediction based on a simulated navigator (model configuration N0) yields better results than spatial prediction based on the respiratory signal (R0 and RS0). The other configurations are all based on the respiratory signal, with differences in forward-prediction time. It is clear that, for the Dice coefficient, the median surface distance and the blood vessel misalignment, a larger time step in prediction yields a lower accuracy in the results. This holds for both the R-configurations and the RS-configurations. Finally, the RS-configurations yielded better results for blood vessel prediction than the R-configurations. The Dice coefficient and the surface distance were similar for both configurations.

We analyzed the errors in the extended Kalman filter separately by computing the absolute differences of the normalized respiratory amplitudes predicted by the Kalman filter and the actual values obtained from the respiratory signal. This was done for the time steps of 300, 600, 900 and 1200 ms. These results are show in Figure 6. The results
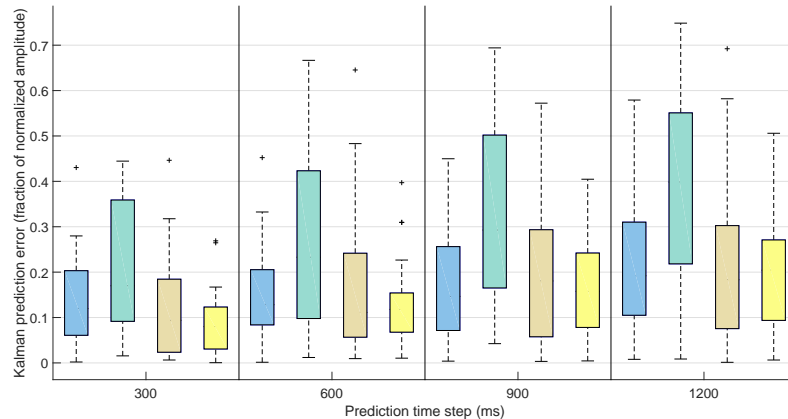
Figure 6: The Kalman prediction errors, as fractions of the normalized respiratory amplitudes, plotted against the prediction time step. The different colors represent the 4 different subjects.

Table 3: Kalman prediction errors (fractions of the normalized amplitudes) in extreme situations.

| Time step (ms)	Frequency change	Sudden stop	Deep inhalation
300	0.69 [0.63-0.75]	0.41 [0.26-0.49]	0.74 [0.59-1.00]
600	0.95 [0.87-1.08]	0.53 [0.26-0.69]	0.95 [0.87-1.02]
900	1.09 [1.01-1.23]	0.62 [0.31-0.86]	1.06 [0.91-1.23]
1200	1.06 [1.03-1.08]	0.71 [0.30-0.92]	1.19 [0.91-1.34]

show that the Kalman error increases with a larger time step for all subjects.

In addition, the performance of the Kalman filter was tested in extreme situations (see Table 3). It was found that the errors produced by the Kalman filter increased with a larger prediction time step. Also, they could be higher than the normalized amplitude. For sudden frequency changes and deep inhalations, the errors were caused by a relatively small phase shift, which could result in large amplitude errors. For the sudden stops, these errors were caused by the sudden change of the amplitude. It was found that the Kalman filter was able to correct these errors in 3.6 to 4.5 seconds. Figure 7 shows the Kalman filter errors for a sudden stop.

The correlation coefficients between the Kalman prediction errors and the errors in our final results were computed. These results are presented in Table 4. The correlations that were calculated for each subject separately show that correlations were found for subjects 2 to 4. However, for subject 1, no significant correlations were found for all measures. The bottom row shows the correlation for all subjects. As expected, a negative correlation exists between the Kalman prediction error and the Dice coefficient, whereas the correlation is positive for the median surface distance and the vessel misalignment. All these correlations are significant, as indicated by the
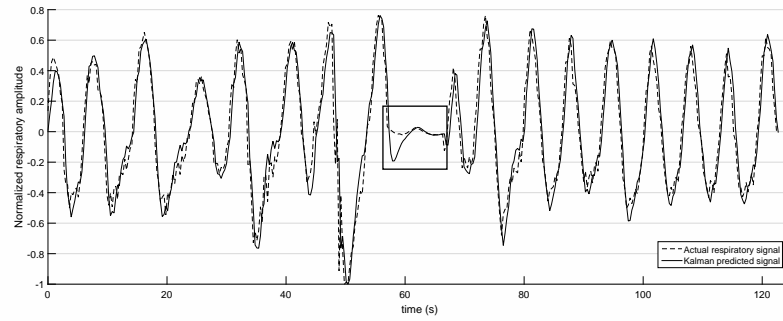
(a) Kalman prediction compared to the original signal during a sudden stop.
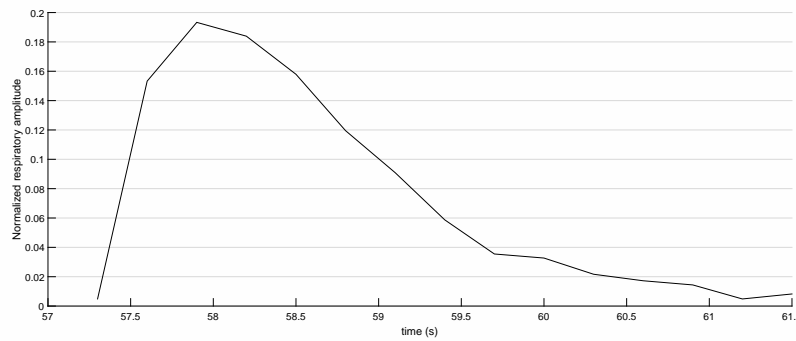
(b) The Kalman error during the sudden stop.

Figure 7: Sudden stop during Kalman prediction.

Table 4: The correlation coefficients r and their significances (P-values) between the Kalman prediction error and model outcomes.

Subject	Dice coefficient		Median surface distance		Vessel misalignment	
	r	P-value	r	P-value	r	P-value
1	-0.0965	0.1361	0.077	0.2344	-0.0385	0.5527
2	0.4295	3.440E-12	0.4204	1.073E-11	0.4518	1.799E-13
3	-0.4669	2.143E-14	0.4623	4.138E-14	0.5275	1.367E-18
4	-0.2989	2.413E-06	0.3071	1.24E-06	0.3234	3.031E-07
Total	-0.2144	1.898E-11	0.2100	4.952E-11	0.2557	8.64E-16

P-values, for all of which holds: $P \ll 0.05$.

The performance of the model on samples with a small Kalman error (< 0.2) and a large Kalman error (≥ 0.2) is presented in Table 5. This evaluation was performed on the R300-R1200 and RS300-RS1200 configurations only, since the other configurations did not employ Kalman prediction. The differences in the Dice coefficient run from 0 for R300 to 0.016 for R1200. For the median surface distances, the differences between the two groups vary between 0.01 mm for RS300 and 0.61 mm for R1200. The difference in performance is most visible in the vessel misalignment results, which show an average difference of 0.7 mm between both groups.

Table 5: Comparison of averages outcomes for a Kalman error smaller and larger than 20% of normalized amplitude.

Model configuration	multicolumn2c— Dice coefficient		Median surface distance (mm)		Vessel misalignment (mm)	
Kalman error	< 0.2.	≥ 0.2	< 0.2	≥ 0.2	< 0.2	≥ 0.2
R300	0.944 ± 0.026	0.944 ± 0.019	1.97 ± 1.04	2.05 ± 0.90	3.11 ± 1.32	3.94 ± 1.67
R600	0.945 ± 0.027	0.938 ± 0.028	1.93 ± 0.96	2.35 ± 1.26	3.29 ± 1.57	3.96 ± 1.70
R900	0.943 ± 0.028	0.932 ± 0.036	2.03 ± 1.08	2.57 ± 1.67	3.50 ± 1.76	4.11 ± 1.99
R1200	0.945 ± 0.030	0.929 ± 0.035	2.05 ± 1.18	2.66 ± 1.59	3.61 ± 1.60	4.31 ± 2.20
RS300	0.941 ± 0.026	0.942 ± 0.018	2.14 ± 1.07	2.15 ± 0.83	3.14 ± 1.55	3.78 ± 1.58
RS600	0.945 ± 0.027	0.938 ± 0.027	1.96 ± 0.97	2.32 ± 1.24	3.27 ± 1.81	3.77 ± 1.46
RS900	0.943 ± 0.028	0.933 ± 0.036	2.05 ± 1.09	2.48 ± 1.63	3.37 ± 1.71	4.13 ± 2.15
RS1200	0.944 ± 0.031	0.931 ± 0.031	2.14 ± 1.25	2.56 ± 1.40	3.45 ± 1.39	4.11 ± 2.05

Discussion

The main goal of this paper was to investigate the clinical potential and the limitations of the liver motion model we proposed in [6]. For this purpose, different situations were evaluated. Firstly, a comparison was made between building the model based on a simulated navigator signal (as was the case in the aforementioned paper) and based on a respiratory signal. Secondly, we compared the performance of spatial prediction with forward-prediction for 1 to 4 time steps (300 to 1200 ms). Thirdly, we analyzed the influence of removing extreme situations, as could be detected in the respiratory signal that was used during the model building process. Finally, we evaluated the Kalman filter performance for all samples and also separately for extreme situations.

The results for the N0 configuration were in line with the results that were obtained in our previous work [6], with an average Dice coefficient of 0.96, an average median surface distance of 1.3 mm and an average vessel misalignment of 2.9 mm. For the first subject, the Dice coefficient and median surface distance yielded good results compared with the other subjects, while the vessel misalignment was the poorest for this subject. It seemed that the internal liver deformation for this subject was relatively large, which affects the registration accuracy during the building of the model. Small prediction errors may have a larger effect in this subject as well, since the deformation fields between different model states show larger differences, yielding a relatively large error when selecting a wrong state from the look-up table.

For all three measures, the navigator-based model (N0) outperformed the respiratory-based configurations. A difference of 0.013 was found for the average Dice coefficient, whereas the average median surface distance increased with 0.68 mm and the average vessel misalignment increased with 0.51 mm (see Table 2). This should be expected, since the navigator yields a direct measurement of liver displacement, in contrast to the respiratory signal. However, this simulated navigator could only be

calculated offline after the data acquisition and did not yield a measure that could be used during treatment. As mentioned before, a 4D dynamic sequence with a built-in navigator echo signal would probably be the optimal solution, but this is difficult to realize in practice. Using the respiratory signal as an indirect measurement of liver motion is the most practical solution. To translate the respiratory amplitude and phase to the corresponding position in the liver motion cycle, a relationship needs to be established between them. In this work, we analyzed the phase difference between both signals on the part of the breathing signal that was acquired before the start of the scan and used this phase delay to shift the respiratory signal. An additional error of the model is expected in this case.

From Table 2, it is clear that a larger prediction time yields less accurate results. This additional error can be attributed to the Kalman filter. Temporal prediction of more than one time step is achieved by not updating the Kalman filter after every time step, but instead performing the prediction step several times consecutively. Comparing the results for a time step of 0 ms to a time step of 1200 ms, the Dice coefficient decreased with 0.85%, the median surface distance increased with 20.6% and the vessel misalignment increased with 20.4%. This indicates that prediction of almost 1 s ahead is feasible. Even though the errors increased, a mean vessel misalignment of 4 mm could still be obtained. Since an additional margin of 0.5 to 1.0 cm is taken into account in clinical practice [16], this is still a promising result for application in the clinic.

Excluding extreme situations from the model building process (RS-configuration) yielded slightly better results than including all samples (R-configuration). The increase in performance was small however, since the model probably suffered from the fact that a significant part of dynamics were missing, yielding fewer states in the final model due to the grouping process. In a clinical setting, a patient would be sedated, such that no considerable irregularities should be expected in the breathing pattern during treatment. In these experiments, the RS-configuration and the R-configuration had similar performance, indicating that the presence of irregularities does not necessarily have a significant influence, provided that sufficient samples are present to average out their effects.

The performance of the signal prediction by the extended Kalman filter was evaluated separately. The results show that these errors increase with a larger time step. As shown in Figure 6, the median error value for a prediction time step of 300 ms is well below 20% of the normalized amplitude, which is considered a small error, since this corresponds to the average distance between two consecutive model states. Although this error increases with a larger time step, for three of the four subjects, the median Kalman error is still approximately 0.2 for a prediction time step of 1200 ms.

In extreme situations, the average Kalman error was found to vary between 0.41 and 1.09. Such errors typically lead to misalignments of 8 to 20 mm. For a sudden stop in the breathing pattern, the Kalman filter typically needed 3.6 to 4.5 seconds to recalibrate. For changes in frequency and sudden deep inhalations, the errors were mainly caused by small phase differences and were corrected in less than a second. Depending on the

treatment for which the model is used, it should be decided if the change in delivered dose resulting from these errors (either heat or radiation) is acceptable for the patient. It is also possible to stop the treatment as soon as such outliers are detected, since the Kalman filter updates continuously, such that errors can be checked in real-time. It might be feasible to improve the final results by using a more complex Kalman filter implementation, for example by approximating the motion by a combination of harmonic functions with different frequencies, instead of only one.

Comparing the results for samples with a Kalman prediction error larger than 0.2 and a prediction error smaller than 0.2, we observed that a larger Kalman prediction error leads to a larger final error. This is also confirmed by the correlation values (Table 4), which show a strong correlation between the Kalman error and the final errors in the model. The difference in average errors in Table 5 is smaller than the additional error of 4 mm that is expected with a Kalman error of 0.2. This suggests that registration errors and grouping errors are present in all samples, and these decrease the difference between samples with a Kalman error smaller than 0.2 versus samples with an error equal to or larger than 0.2. Overall, the temporal prediction using a Kalman filter based on the respiratory signal adds a vessel misalignment error of 0.3 to 1.1 mm.

Conclusion

In this paper, the practical limitations and the performance in a more applied setting were evaluated of a liver motion model based on registration of dynamic slices. This model was first proposed for spatial prediction in our earlier work [6] and was extended to temporal prediction based on a respiratory signal in this work. An extended Kalman filter based on harmonic motion was implemented to predict the liver translation into the future. The original implementation, using a simulated navigator signal, was compared with an implementation based on the respiratory signal. Forward-prediction was performed for four time intervals of 300, 600, 900 and 1200 ms. In a clinical setting, the necessary prediction time would probably not be more than 1 s, since this is the typical time needed for feedback to the HIFU system and mechanical adjustment of the beam. However, many systems make use of electronical steering, in which different transducer elements are switched on and off throughout the treatment to adjust the HIFU focus. With such a system, only tens of ms [17, 18] would be needed to update the beam configuration. In that case, the prediction time step of 300 ms would already be sufficient. The errors caused by the Kalman prediction were evaluated separately and performance in extreme situations was investigated. The results showed that extreme errors could occur, but that the Kalman filter was also able to correct these errors in a few seconds. On average, a prediction of 300 ms to 1200 ms ahead yields an additional vessel misalignment error of 0.3 to 1.1 mm. This leads to a total average error of 3.3 to 4.1 mm, which could still be acceptable for clinical use.

Acknowledgments

References

[1] Preiswerk F, Arnold P, Fasel B and Cattin P C 2012 A Bayesian framework for estimating respiratory liver motion from sparse measurements *Abdominal Imaging. Computational and Clinical Applications* (*Lecture Notes in Computer Science* vol 7029) ed Yoshida H, Sakas G and Linguraru M G (Berlin Heidelberg: Springer-Verlag) pp 207–14

[2] Samei G, Tanner C and Székely G 2012 Predicting liver motion using exemplar models *Abdominal Imaging. Computational and Clinical Applications* (*Lecture Notes in Computer Science* vol 7601) ed Yoshida H, Hawkes D and Vannier M W (Berlin Heidelberg: Springer-Verlag) pp 147–57

[3] von Siebenthal M, Székely G, Gamper U, Boesiger P, Lomax A and Cattin P 2007 *Phys. Med. Biol.* **52** 1547–64

[4] de Senneville B, Mougenot C and Moonen C 2007 *Magn. Reson. Med.* **57** 319–30

[5] Blackall J M, Ahmad S, Miquel M E, McClelland J R, Landau D B and Hawkes D J 2006 *Physics in Medicine and Biology* **51** 4147

[6] Noorda Y, Bartels L, Viergever M and Pluim J 2016 *J Med Imag* **3** 015002

[7] Kalman R 1960 *J. Basic. Eng.* **82** 35–45

[8] Riadh K 2006 *Different Aspects of Nonlinear Stochastic Filtering Theory* Ph.D. thesis Universität Mannheim

[9] Müller 2008 Deconvolving oscillatory transients with a Kalman filter Report University of Applied Sciences, Switzerland

[10] Lewis F, Xie L and Popa D 2007 *Optimal and Robust Estimation: With an Introduction to Stochastic Control Theory, Second Edition* (Boca Raton: CRC Press)

[11] Wigren T, Abd-Elrady E and Söderström T 2003 Harmonic signal analysis with Kalman filters using periodic orbits of nonlinear ODEs *Acoustics, Speech, and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference On* vol 6 (ICASSP) pp VI–669–72

[12] Roujol S, de Senneville B D, Hey S, Moonen C and Ries M 2012 *IEEE transactions on medical imaging* **31** 533–542

[13] Wood M L and Henkelman M 1985 *Med. Phys.* **12** 143–51

[14] George R, S V S, Chung T, Ramakrishnan V and Keall P J 2005 *Med. Phys.* **32** 2850–61

[15] Malcolm J, Rathi Y, Yezzi A and Tannenbaum A 2008 Fast approximate surface evolution in arbitrary dimension *Proc. SPIE* vol 6914 pp 69144C–69144C–9

[16] Wells S A, Hinshaw J L, Lubner M G, Ziemlewicz T J, Brace C L and Lee F T 2015 *Radiologic Clinics of North America* **53** 933–971

[17] Aubry J F, Pauly K B, Moonen C, Haar G, Ries M, Salomir R, Sokka S, Sekins K M, Shapira Y, Ye F *et al.* 2013 *Journal of therapeutic ultrasound* **1** 1

[18] Pernot M, Tanter M and Fink M 2004 *Ultrasound in medicine & biology* **30** 1239–1249