Multi-modal respiratory motion prediction using sequential forward selection method

R. Dürichen[1,2], T. Wissel[1,2], F. Ernst[1], A. Schweikard[1]
[1] University of Lübeck, Institute for Robotics and Cognitive Systems, Lübeck, Germany
[2] University of Lübeck, Graduate School for Computing in Medicine and Life Sciences, Lübeck, Germany

contact: duerichen@rob.uni-luebeck.de

Abstract:

In robotic radiotherapy, systematic latencies have to be compensated by prediction of external optical surrogates. We investigate possibilities to increase the prediction accuracy using multi-modal sensors. The measurement setup includes position, acceleration, strain and flow sensors. To select the most relevant and least redundant information from the sensors and to limit the size of the feature set, a sequential forward selection (SFS) method is proposed. The method is evaluated with three prediction algorithms – the least means square (LMS) algorithm, a wavelet-based LMS algorithm (wLMS) and an algorithm using relevance vector machines (RVM). We show that multi-modal inputs can easily be integrated into general algorithms. The relative root mean square error (RMS_{rel}) of the best predictor, RVM, could be decreased from 60.5 % to 49.0 %. Furthermore, the results indicate that more complex algorithms can efficiently use different modalities like acceleration which are less correlated with the optical sensor to be predicted.

Keywords: radiosurgery, respiratory motion prediction, feature selection

1 Problem

In extracranial radiotherapy, many tumors are constantly moving due to respiration of the patient. The movement has to be compensated to allow precise irradiation of the tumor while sparing critical surrounding structures. Modern technical systems, like multileaf collimators, robotic patient couches or the CyberKnife, allow active motion compensation based on the correlation between internal and external surrogates [1]. The internal motion is measured via stereoscopic X-ray cameras and the external surface of the patient via multiple optical markers. Each technical system has to compensate systematic latencies, due to image acquisition, data processing and mechanical limitations. The resulting systematic error can be decreased by time series prediction of the external surrogates. The latency of the current CyberKnife system is 115.5 ms.

Several studies investigated the correlation between external and internal movements. Thereby, the focus was mainly on external optical markers. Even though the correlation was in general high, Ahn *et al.* [2] concluded that the correlation depends significantly on the marker placement, the tumor position and the breathing characteristics.

We investigate how the prediction and correlation accuracy can be increased using a multi-modal sensor setup. The setup includes ultrasound, acceleration, flow, strain and optical tracking sensors. Preliminary results indicate that the correlation between these sensors depends strongly on the breathing characteristics, the modality and the position of the sensor [3]. Here, we want to present the first results to use this setup for multi-modal sensor prediction. In this study, a sequential forward selection (SFS) method is used to select the best input signals for the prediction. The potential of this approach is evaluated with three different prediction algorithms – a least mean square (LMS) algorithm, a wavelet based LMS (wLMS) algorithm and a relevance vector machine (RVM). Unfortunately, the concrete implementation of the CyberKnife algorithm is unknown and could not been used for this study. However studies on clinical data have shown that wLMS can outperform this algorithm [4].

2 Methods

2.1 Data set

The data set consists of 18 subjects (11 male / 7 female, aged: 21 – 34 years). Each subject was asked to breathe normally for 20 minutes. The first and the last minutes of each recording were discarded to eliminate potential irregular breathing periods due to initialization or finalization of the measurement. Figure 1.a shows the placement of all six external sensors. As it is clinical standard for CyberKnife treatments, three optical markers (OM) were used. They were placed along the median line of the thorax and abdomen, starting with OM1 close to the areolas mammae, OM2 at the bottom end of the sternum and OM3 above the navel. The position was measured with an accuTrack 250 system (Atracsys LLC, Switzerland). The enlargement of the thorax was measured with a strain belt sensor (SleepSense Piezo

Effort Sensor) which was placed between OM1 and OM2. The air flow was measured indirectly with a thermistor (SleepSense Airflow Thermistor). The strain belt and the flow sensor were connected to a g.tec USB amplifier (g.tec medical engineering GmbH, Austria). An acceleration (ACC) sensor was placed between OM1 and the strain belt, which measured the acceleration in all three directions with a range of ±2 g. All sensors were synchronized via strobe values and downsampled to a sample frequency of f_s = 26 Hz, which is used in the current CyberKnife system. The 3D position and acceleration signals were reduced to their first principal component. More details about the setup are given in [3].

2.2 Time series prediction

To compensate the latency error, the prospective true signal amplitude of an external surrogate $y_{t+\delta}$ can be predicted. Here, t is the index of the current time step and δ is the prediction horizon, which was fixed to δ = 3 according to the latency of the CyberKnife system. The output of a prediction algorithm is denoted by . In general, the prediction is based on an input vector $\boldsymbol{y}_t = \{y_t, y_{t-1}, \ldots, y_{t-p+1}\}$ of the current and previously observed data points, where p is the number of features per input vector. In recent years, several prediction algorithms have been proposed, which try to learn the underlying function $f(\boldsymbol{y}_t)$. As many of these algorithms are general prediction algorithms, they are not restricted to a certain signal modality or to a certain size of $\boldsymbol{y}_t$ like e.g. Kalman Filters. Therefore, to predict for example the signal of the second optical marker , $\boldsymbol{y}_t$ can be extended by the input vectors of OM1 and ACC to $\boldsymbol{y}_t^{OM2,OM1,ACC} = \{\boldsymbol{y}_t^{OM2}, \boldsymbol{y}_t^{OM1}, \boldsymbol{y}_t^{ACC}\}$. Assuming a constant p, the dimension of the input vector would increases by a factor of three.

To evaluate our approach, we perform multi-modal prediction with three prediction algorithms. The first algorithm is the classical LMS algorithm which performs predictions according to $= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}_t$. The weight vector $\boldsymbol{w}$ is adaptively learned at each time step t by minimizing the current prediction error $e_t = y_t - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}_t$. To increase the prediction accuracy and stability, the error of M points can be considered simultaneously at each time step t. To account for information beyond the signal history M, an averaging parameter $\mu \in [0, 1]$ is added.

The second algorithm is the wLMS algorithm [5]. The algorithm decomposes a signal into $J+1$ scales using an à trous wavelet. A separate LMS predictor can be applied to each scale. The predicted point is the sum of all predicted points per scale. Ernst *el al.* [6] showed on a dataset of 304 motion traces that, using J = 3, M = 193 and μ = 0.0204, the wLMS algorithm can outperform support vector regression, LMS and Kalman Filter.

Finally, we used a predictor based on relevance vector machines (RVM). The RVM is a probabilistic approach using Bayesian inference. In [7], it has been shown that the RVM can outperform wLMS on average. Here, a multi-modal RVM with a linear function was implemented, to reduce the computational complexity.

To compare the results of N predicted points, irrespective of the patient-specific respiration amplitude, all results were evaluated with respect to the relative root mean square RMS_{rel} = . A RMS_{rel} = 100 % means that the prediction accuracy could not be improved compared to no prediction and for values above 100 % that the RMS error of the specific predictor is higher than for no prediction.

2.3 Sequential forward selection

As in many classification and regression problems, the selection of the optimal feature set S_{opt} among all possible features F becomes essential to prevent overfitting and to reduce computation time, once the amount of input features increases. Feature selection methods can be divided into wrapper and filter methods [8]. Filter methods use information criteria like correlation coefficient or mutual information to estimate the relevance of a feature to the output and the redundancy between features.

In contrast, wrapper methods use a specific prediction algorithm and evaluate the relevance and redundancy of a feature depending on the prediction accuracy. Even though wrapper methods are often computationally expensive, they find the optimal features set for the specific algorithm. They can roughly be divided into forward selection and back-

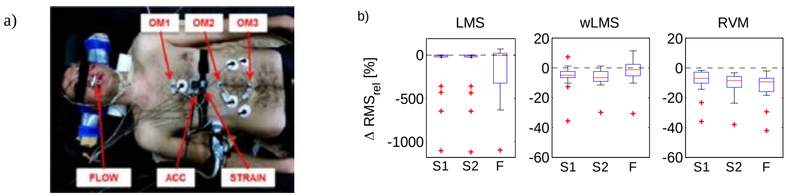

Figure 2: a) Placement of the optical markers (OM1-3), strain belt, flow and acceleration (ACC) sensors; b) Rel. RMS error difference for LMS, wLMS and RVM, defined as the difference between RMS_{rel} using S_1, S_2 or F and S_0.

ward elimination [8].
Here, the features correspond to the used sensors. The complete feature set is F = {OM1, OM2, OM3, STRAIN, FLOW, ACC}. For the prediction of an external optical sensor, a sequential forward selection (SFS) method is very applicable compared to backward elimination, as within all possible features the optical sensor itself will have the highest relevance to the output. Therefore, the initial feature set S_0 of the SFS method is not the empty set. Assuming that OM2 shall be predicted, the initial feature set is S_0 = {OM2}. With this feature set a prediction error $RMS_{rel\text{-}S0}$ can be achieved. In the next step, S_0 will be expanded to S_1 by one of the remaining features F\{OM2}. The feature that has the lowest $RMS_{rel\text{-}S1}$ is selected. This procedure can be repeated up to a certain size d of the feature set S or until no features are found which further decrease RMS_{rel}.

2.4 Experiment

In our experiment, OM2 was exemplarily chosen to be the prediction target. We divided each measurement into a training set (t_{train} = 1 min) and a test set (t_{test} = 17 min). The learning factor μ and the number of features per input vector p were optimized on the training set using exhaustive grid search. All wavelet scales of wLMS use the same p. The history length was set to $M_{LMS} = 1$, $M_{wLMS} = 193$ [5], $M_{RVM} = 1000$ [7] and $J = 3$. The initial feature set was set to S_0 = {OM2}.The maximum size of the feature set was set to $d = 2$, leading to two subset S_1 and S_2. For comparison, the RMS_{rel} was also computed for the complete features set F.

3 Results

Fig. 1.b shows a boxplot of the rel. RMS error differences ΔRMS_{rel}, which is defined as the difference between RMS_{rel} using S_1, S_2 or F and RMS_{rel} using S_0. The median ΔRMS_{rel} is negative for all prediction algorithms and all feature sets, indicating that, in general, prediction accuracy is increased by using multiple sensors. The mean RMS_{rel} (and standard deviation) for all feature sets and all algorithms is listed in table 1. On average, LMS has the highest prediction error, followed by wLMS and RVM across all feature sets. Using the complete set F, the error decreases on average by 154.4 percentage points (pp) for LMS, by 2.9 pp for wLMS and by 12.9 pp for RVM compared to the results using S_0. Using the SFS method with S_1 and S_2, the mean RMS_{rel} is smaller than the mean RMS_{rel} using F for LMS and wLMS.
Table 2 shows the composition of the feature sets S_1 and S_2. The algorithms use different features to improve the prediction accuracy. While the LMS algorithm uses mostly information from the optical sensors, wLMS and RVM use optical, strain and acceleration information. The flow sensor is the least used sensor. Furthermore, table 2 shows n_{S1} and n_{S2}, the number of cases for which a feature set S_1 or S_2 was actually computed – meaning that there was another feature which decreased the RMS_{rel}. RVM could find an improved subset S_1 and S_2 for all cases. In contrast, n_{S1} is 12 and 17 and n_{S2} is 6 and 11 for LMS and wLMS, respectively.

4 Discussion

The results of the experiment indicate that, in general, prediction accuracy can be increased using multi-modal sensors. However, using simply all features does not necessarily lead to the best RMS_{rel} as shown in fig. 1.b for the wLMS algorithm. Using all features can lead to overfitting of the data. Even though not analyzed here, using the complete data set F also results in a strong increase of the computation time. Using a sequential forward selection of the features, an improved predictor-specific feature set can be found that uses multiple features and prevents overfitting of the data. The computational requirements and the size of the feature set can be limited with the factor d.
Comparing the three evaluated prediction algorithms reveals that RVM has the highest prediction accuracy, being in agreement with [7]. Furthermore, RVM can use the complete feature set F without overfitting of the data. However, the RMS_{rel} difference between S_2 and F is only 1.4 pp while the size of the features set is doubled. The high error of LMS was expected, as LMS is implemented in its simplest version with $M = 1$. With increasing M, the error would decrease. LMS serves here more as an example of a less complex algorithm. On average, the prediction performances of wLMS and RVM are worse compared to other publications [6], [7]. One reason is that even though, all subjects were asked to remain still, several movement artifacts could be identified in the measurements. The artifacts lead to the strong outliers.

	Mean RMS_{rel} (standard deviation) in %		
	LMS	**wLMS**	**RVM**
S_0	251.6 (314.6)	63.3 (14.2)	60.5 (19.8)
SFS, S_1	104.3 (120.4)	57.1 (12.4)	51.3 (12.7)
SFS, S_2	103.0 (120.6)	58.3 (17.0)	49.0 (12.8)
F	97.2 (14.1)	60.4 (12.4)	47.6 (12.4)

Table 1: Mean RMS_{rel} and standard deviation (in parentheses) of the LMS, wLMS and RVM algorithms using the initial feature set S_0 with data only of OM2, the feature sets S_1 and S_2 determined by SFS and the set of all F.

		OM1	OM3	Flow	Strain	ACC	n_{S1} / n_{S2}
LMS	S_1	2	10	0	0	0	12
	S_2	5	1	0	0	0	6
wLMS	S_1	0	3	0	10	4	17
	S_2	0	7	0	3	1	11
RVM	S_1	0	3	0	12	3	18
	S_2	1	10	1	2	4	18

Table 2: Distribution of the features which were selected by SFS to be added to the subset S_1 and S_2 and number of cases n_{s1} / n_{s2} in which a feature set S_1 and S_2 could be computed for LMS, wLMS and RVM over all 18 subjects.

Using SFS or the complete data set *F*, the RMS_{rel} could be decreased strongly (crosses in fig. 2). Even though a real treatment would be interrupted at the occurrence of a strong motion artifact, these results indicate that also the effect of smaller motion artifacts could be better compensated using multi-modal sensors.

Table 2 illustrates that there is not only one optimal feature which could be added to decrease RMS_{rel} in general. The features selected for S_1 and S_2 strongly depend on the prediction algorithm. Thereby, the complexity of the algorithm seems to be correlated to the selected features. The less complex LMS model uses basically only optical features. As shown in [3], these features have a high correlation to OM2 as they are measuring the same signal modality. More complex models can incorporate features based on different modalities which are less correlated to OM2 [3]. These features are potentially more relevant for the output and less redundant to the already selected features. This results also in the different numbers of n_{S1} and n_{S2}. As the optical features are highly correlated to each other, they contain redundant information. As in the case of LMS, the expansion of S_1 by another optical feature does not decrease the RMS_{rel} as much of this information is already contained in the first two optical features. Consequently, the SFS method stops and n_{S2} is small for LMS. In contrast, RVM and wLMS can use features which are less redundant to each other leading to higher values of n_{S2}. There is a strong variation of the added features for S_1 and S_2 in case of wLMS and RVM. This variation is most likely due to the heterogeneous subject group (male / female) and different breathing patterns which each subject has. The flow sensor is only selected in two cases. A probable reason could be the dead times of the airways and the thermistor, which lead to temporal delays between the mechanical movement of the torso / abdomen and the temperature difference of the air flow in front of the nose / mouth. Latter could be decreased by using an aeroplethysmograph.

As mentioned in section 2, the SFS method itself can be computationally expensive depending on the algorithm and the training set. This could impede the use of this method in real time applications. An alternative feature selection method could be based on filter methods. However, it has to be further investigated how these methods could be used efficiently to find the optimal feature set.

5 Conclusion and summary

In this paper, we used a multi-modal sensor setup to increase the prediction accuracy in respiratory motion prediction. We demonstrated that the most relevant and least redundant sensors can be selected by an SFS method, which lead to a decrease in RMS_{rel}. We show that all evaluated algorithms can be easily adapted to use multi-modal inputs. As all sensors are relatively inexpensive, this technique could be easily integrated in treatment rooms.

6 References

[1] A. Schweikard, G. Glosser, M. Bodduluri, M. J. Murphy, and J. R. Adler, "Robotic Motion Compensation for Respiratory Movement during Radiosurgery," *Comput. Aided Surg.*, no. 5, pp. 263–277, 2000.

[2] S. Ahn, B. Yi, Y. Suh, J. Kim, S. Lee, S. Shin, S. Shin, and E. Choi, "A Feasibility Study on the Prediction of Tumour Location in the Lung from Skin Motion," *Br. J. Radiol.*, vol. 77, no. 919, pp. 588–596, Jul. 2004.

[3] R. Dürichen, L. Davenport, R. Bruder, T. Wissel, A. Schweikard, and F. Ernst, "Evaluation of the Potential of Multi-modal Sensors for Respiratory Motion Prediction and Correlation," in *IEEE Engineering in Medicine and Biology Society - EMBC 2013*, Osaka, Japan, 2013.

[4] O. Blanck, R. Dürichen, Ernst, Floris, Dunst, Jürgen, Rades, Dirk, Hildebrandt, Guido, and Schweikard, Achim, "Evaluation of a wavelet-based least mean square motion prediction algorithm for lung and liver patients," no. ESTRO 31, Barcelona Spain, 2012.

[5] F. Ernst, A. Schlaefer, and A. Schweikard, "Prediction of Respiratory Motion with Wavelet-Based Multiscale Autoregression," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*, vol. 4792, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 668–675.

[6] F. Ernst, R. Dürichen, A. Schlaefer, and A. Schweikard, "Evaluating and comparing algorithms for respiratory motion prediction," *Phys. Med. Biol.*, vol. 58, no. 11, pp. 3911–3929, Jun. 2013.

[7] R. Dürichen, T. Wissel, F. Ernst, and A. Schweikard, “Respiratory Motion Compensation with Relevance Vector Machines,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013*, Nagoya, Japan, 2013.

[8] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J Mach Learn Res*, vol. 3, pp. 1157–1182, März 2003.