SPEEDING-UP IMAGE REGISTRATION FOR REPETITIVE MOTION SCENARIOS

Valeria De Luca, Christine Tanner, Gábor Székely

Computer Vision Laboratory, ETH Zürich,
Sternwarstrasse 7, 8092 Zürich, Switzerland

ABSTRACT

We propose a novel approach for real-time image registration for image sequences of organs subject to repetitive movement, such as breathing. The method exploits the redundancy within the images and consists of a training and an application phase. During training, the images are registered and then the relationship between the image appearance and the spatial transformation is learned by employing dimensionality reduction to the images and storage of the corresponding displacements. For each image in the application phase, the most similar images in the training set are selected for predicting the associated displacements. Registration and update of the training data is only performed for outliers. The method is assessed on 2D sequences (4 MRI, 1 ultrasound) of the liver during free breathing. The performance is evaluated on manually selected landmarks, such as vessel centers and the distal point of the inferior segment. The proposed algorithm is real-time (9 ms per frame) and the prediction error is on average 1.2 mm for both MRI and ultrasound.

***Index Terms*—** Respiratory motion, tracking, image registration, magnetic resonance imaging, ultrasound imaging

1. INTRODUCTION

Recent technical developments in radiation therapy, such as intensity-modulated radiation therapy and high intensity focused ultrasound (HIFU), provide the capability to deposit highly conformal radiation dose distributions into the tissues. Therefore compensation for any movement in the treatment region, like respiratory motion in the abdomen, is vital [1, 2]. This not only requires an accuracy in the range of millimeters but also real-time capability.

Unfortunately, observation of the target (e.g. the tumor) motion in real-time during therapy is often impossible. Instead, methods have been proposed based on surrogate measures of the target motion and a model which relates target and surrogate motion [1]. Studies of liver motion showed that an indirect strategy, based on tracking a few surrogate markers inside the liver, allowed to accurately estimate drifts of the organ by a statistical motion model [3, 4].

Common surrogates include external signals such as recordings from a breathing bellow or a spirometer, and internal measurements from fast imaging. The latter could be achieved by fluoroscopy, ultrasound (US) or cine MRI. Studies have shown that internal surrogates correlate better with the target motion than external respiratory signals in the abdomen during breathing [1].

Several techniques were proposed to handle real-time respiratory organ movement in image-guided applications. Most of the current tracking systems are based on fluoroscopy images [5] with the major disadvantages of higher dose to the patient and the need to implant fiducial markers. Ries *et al.* [6] proposed tracking of 2D MRIs based on a Kalman predictor combined with a compensation for through-plane motion. Experiments carried out on images of the kidney under regular respiratory motion due to mechanical ventilation show tracking precision of $\pm$ 1.1 mm and processing time of 40 $\pm$ 8 ms. US imaging represents the best modality for image-guidance as it is real-time, cheap and non-ionizing. US image sequences have been tracked using speckle tracking algorithms [7] with an accuracy of 1.7 mm. Other methods are based on feature extraction and registration [8]. Volumes were registered with an error of 1.3 mm, yet missing real-time performance. A 2D real-time US tracking method based on active contours and conditional density propagation was proposed in [9] and evaluated on a short sequence. Results showed an accuracy of 1.6 mm and a processing time of 5 ms per frame.

While feature-based tracking methods are potentially fast, they might suffer from bad initialization, feature disappearances and feature location inaccuracies. Intensity-based methods do not require feature extraction and hence can work in regions with weak features. They usually need no preprocessing, but are generally computationally expensive. In this study, we investigate how intensity-based image registration employed for tracking respiratory motion can by accelerated by exploiting the redundancy in the data.

We propose a novel approach for real-time image registration for image sequences of organs subject to repetitive movement, such as breathing. During an initial training phase the images are registered and we learn the relationship between the image appearance and the displacements. For each image in the real-time application phase, the most similar images in the training set are selected for predicting the associated displacements. We incorporate a mechanism to cope with unseen variations in the images during the application phase, which allows for tracking of non-periodic motions. Although the presented framework is applicable to any image modality and any organ subject to repetitive movement, we asses its performance for real-time tracking of landmarks in 2D MRI and US sequences of the liver of volunteers under free breathing.

2. MATERIAL

To explore the validity of the method independently from the image modality, we acquired 4 MR sequences and 1 US sequence of the liver of healthy volunteers during free breathing. For MRI, we use the MR navigator slices, previously acquired and registered [10] to generate 4DMRIs. Such 2D images could be employed as surrogates during therapy where the patient is in the MR scanner (e.g. HIFU) [11]. The sequence consists of sagittal slices acquired at the same location at a frequency of 2.8-2.3 Hz over approximately 1 hour. The MRIs have a spatial resolution of 1.4 mm. For US, we acquired 2D images in the longitudinal plane at the midline of the abdomen to capture the movement of the cross section of the left branch of

We thank the Swiss National Science Foundation - Project CRSII2 127549 for funding and the Universitätsspital Zürich, Dr. Orçun Göksel and Dirk Boye (ETH Zürich) for their support in image acquisition.

the hepatic portal vein of the liver. The US sequence has a temporal resolution of 29 Hz and a spatial resolution of 0.16 mm. Images were acquired over a period of approximately 4 min. Images examples are depicted in Fig. 1 for both modalities.

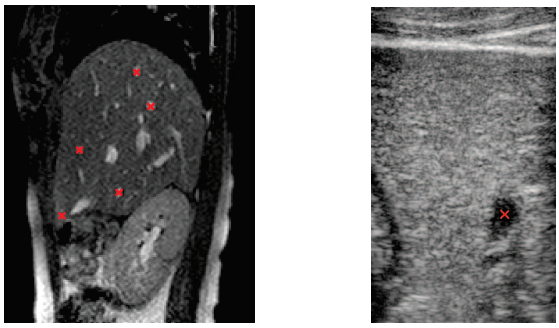


Fig. 1. Examples of a 2D MRI (left) and 2D US image (right) of the liver in exhalation phase. The tracked landmarks are depicted in red.

3. METHOD

3.1. Training Phase

During the training phase we acquire image sequences characterized by D pixels at a frequency rate f Hz resulting in a temporal sequence of T images $\vec{I}(t_i)$, with $t_i = t_0 + i/f$ for $0 \leq i \leq T - 1$.

Image Registration. The images $\vec{I}(t_i)$ are registered to a typical exhalation image (say $\vec{I}(t_0)$) in order to obtain spatial correspondence at each time t_i. This enables the extraction of trajectories for individual points. For MRI, the motion vectors of points within the liver are estimated by intensity-based non-rigid registration using normalized cross correlation as image similarity measure and a B-spline transformation model with 15 mm control point spacing [10, 12]. The 2D US images are registered by optimizing the parameters of a local affine transformation with respect to normalized cross correlation. The transformation is defined for a region enclosing the cross section of a main vessel. To improve accuracy and robustness, we choose a multi-resolution approach with 3 levels. The registration result from the previous image pair ($\vec{I}(t_{i-1})$ to $\vec{I}(t_0)$) is used as an initial estimation.

PCA. In order to store the image appearance efficiently, we use a preprocessing step to reduce the dimensionality of the data. A large number of dimensionality reduction methods have been proposed and compared [13]. On real data, the performance of linear methods was not inferior to that of nonlinear ones. As linear methods are computationally less expensive, we choose PCA for embedding $\vec{I}(t_i) \in \mathbb{R}^D$ into a low-dimensional data representation $\vec{S}(t_i) = [s_1(t_i); \ldots; s_P(t_i)] \in \mathbb{R}^P$,with $P \ll D$.

Each image $\vec{I}(t_i)$ is reshaped into a single column vector $\vec{x}(t_i) \in \mathbb{R}^D$. We calculate the mean with respect to time ($\bar{\vec{x}}$) and the covariance matrix $\vec{C} = \hat{\vec{X}}^T \hat{\vec{X}}$, where the $(i+1)$-th column of $\hat{\vec{X}}$ is equal to $\hat{\vec{x}} = \vec{x}(t_i) - \bar{\vec{x}}$. We then solve the eigenproblem $\vec{C}\vec{w}_j = \lambda_j \vec{w}_j, \forall j \in [1, \ldots, T]$ and rearrange it such that the eigenvalues λ_j are sorted ($\lambda_j \geq \lambda_{j+1}$), with $\vec{w}_j$ being the corresponding eigenvector. For each j-th eigendirection we calculate the data projection $p_j = (\vec{w}_j)^T \hat{\vec{x}}$, so that the original image sequence can be reconstructed by $\vec{x} = \sum_{j=1}^{T} p_j \vec{w}_j + \bar{\vec{x}}$. The most reconstructive information is captured by the eigenvectors associated with the largest eigenvalues and data reduction is achieved by considering only the first P eigenvectors, i.e. $s_j(t_i) = p_j(t_i)$ for $j \in [1, \ldots, P]$. Selection of P is based on the energy accumulated in the first P eigenvectors (calculated by the normalized cumulative sum of the eigenvalues λ_1 to λ_P) as this describes the reconstruction accuracy. We choose P by considering the 95th percentile of the energy.

The low-dimensional representation $\vec{S}$ of the images and the corresponding registration results (e.g. spatial locations for points of interest) are stored at each time step t_i.

3.2. Real-time Application Phase

3.2.1. Displacement Prediction

During a real-time procedure, new images are continuously acquired. Given the current image $\vec{I}(t^*)$, we want to find the most similar image in the training database for obtaining a prediction of the current displacements. For this, we first project the new image into the PCA space:

$$\vec{S}(t^*) = \sum_{j=1}^{P} p_j \vec{w}_j \left(\vec{x}(t^*) - \bar{\vec{x}}\right) \tag{1}$$

Then, we determine the *K*-nearest neighbors (*K*-NN) in the training database to $\vec{S}(t^*)$. In detail, we find t_{m_k} with the smallest distances

$$d_k = \sqrt{\sum_{j=1}^{P} [s_j(t^*) - s_j(t_{m_k})]^2}, \tag{2}$$

$1 \leq k \leq K$. The distances are sorted so that $d_k \leq d_{k+1}$. Having determined the *K*-NN PCA representations from the training set, we retrieve the K associated positions $\hat{\mathbf{P}}(t_{m_k})$ per point of interest $\mathbf{P}(t_0)$. The prediction of $\mathbf{P}(t^*)$ is then obtained by linearly interpolating $\hat{\mathbf{P}}(t_{m_k})$:

$$\bar{\mathbf{P}}(t^*) = \sum_{k=1}^{K} \left(1 - \frac{d_k}{\sum_k d_k}\right) \hat{\mathbf{P}}(t_{m_k}). \tag{3}$$

3.2.2. Outliers

Completely repetitive motion is very unlikely. Therefore a mechanism to cope with previously unseen variations (outliers) and adapt to these (e.g. drift of exhale position) is incorporated. In order to detect such outliers we establish a threshold on the image similarity based on the statistics of the training data. Specifically, for each training image we find its nearest neighbor to the remaining training samples and calculate the minimum Euclidean distance d_1, similar to (2). The threshold θ is then set to the 95th percentile of the distribution of the minimum distances. During the application phase, if $d_K(t^*) > \theta$ then $I(t^*)$ is categorized as outliers and registration is computed as described above in 3.1. Finally we update the training database by adding the registration result and the PCA description of the outlier, and by recalculating θ.

3.2.3. Evaluation

We evaluate the prediction on manually selected landmarks defined on the reference image $\vec{I}(t_0)$. Specifically, for each MR navigator

sequence we annotate vessel centers and the distal point of the inferior subsegment of the posterior segment. For the US sequence, we place the landmark in the center of the tracked vessel, see Fig. 1.

We evaluate the performance by two measures. The prediction error is employed to assess how well the registration result is predicted. The registration error states the difference from manually annotated landmarks, which we call ground truth. In both cases, the error is calculated by the Euclidean distance between the landmark position $\hat{\mathbf{P}}$ and the corresponding location obtained from our method $\bar{\mathbf{P}}$:

$$\bar{E}_{t^*} = \left\| \hat{\mathbf{P}}(t^*) - \bar{\mathbf{P}}(t^*) \right\|. \quad (4)$$

For the prediction error, $\hat{\mathbf{P}}$ is computed by the image registration, while for the registration error it is defined by the manual annotation. We summarize the results by the mean and standard deviation (SD) and the 95th percentile of $\bar{E}$ for all t^* and all landmarks of a sequence. To calculate our results we consider all the landmarks in a single distribution.

3.3. Computational Complexity

The aim of the presented method is to speed-up image registration. We therefore analyze the computational complexity of the proposed algorithm compared to standard image registration. For each image during the application phase, the complexity of our algorithm is:

- Projection of the new image into the training PCA space: $O(PD)$, being P the number of selected eigenvectors and D the image dimension
- K-NN search: let $T_{tr} \leq T$ be the size of the search space (i.e. the number of training images), then the K nearest neighbors can be found in $O(KT_{tr}logT_{tr})$
- Prediction via interpolation: for each 2D point, the complexity is $O(2K)$

For M points, this results in a total complexity of $O(PD + KT_{tr}logT_{tr} + 2KM)$. With $K, M \ll P \ll D$ and $T_{tr} \ll D$, the main complexity is linear in D and can be approximated by $O(PD)$.

For image registration, the complexity per image is $O(IEVR)$, where:

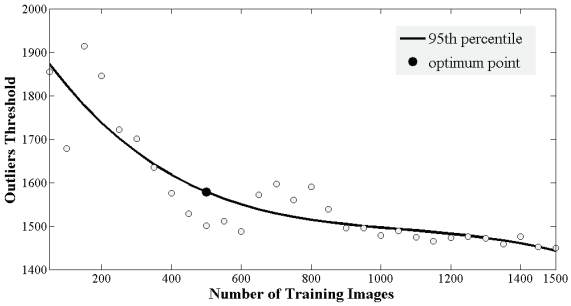
- I is the average number of iterations
- E is the number of image similarity calculations per iteration and transformation parameter. Assuming numerical gradient calculation, $E = 2$
- V is the number of displacements which needs to be calculated for the image similarity per registration parameters. Assuming that the fixed and moving images have the same size N, then $V = N^n$ for affine registration, where n is the image dimension ($n = 2$ for 2D). In the B-spline registration, every control point affects a region of $\pm\, 2\delta s$ in each direction, where δs is the grid spacing. Therefore $V = \left(\frac{4\delta s}{\delta i}\right)^n$, where δi is the spatial image resolution
- R is the number of registration parameters. It is equal to $(n+1)n$ for affine and and Cn for B-spline registration. In addition, $C = \left(N\frac{\delta i}{\delta s} + 1\right)^n$

For 2D images, the complexity can be approximated by $O(12IN^2)$ for the affine registration and by $O(64IN^2)$ for the B-spline registration. With $D \approx N$, the complexity of our method is linear, while the registration reaches quadratic growth.

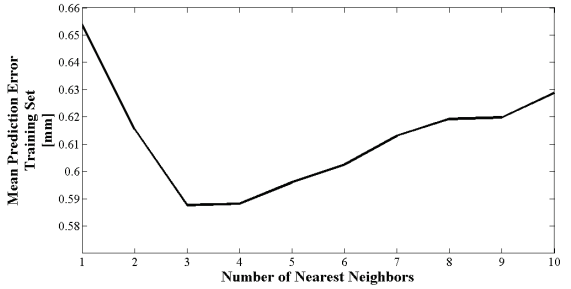
4. RESULTS

We first measure the accuracy of the registration with respect to manually selected landmarks, which serve as our ground truth. For each dataset, we annotate approximately 10-15% of the images from the application phase, as described in Sec. 3.2.3. As a baseline, we also evaluate the motion magnitude of the landmarks defined by $||\hat{\mathbf{P}}(t^*) - \hat{\mathbf{P}}(t_0)||$.

The proposed method has 2 parameters, namely the number of training images (T_{tr}) and the number of nearest neighbors (K). In order to get a good trade-off between the size and the information content of the training set, we apply the L-curve method to determine T_{tr}. Specifically, we calculate the threshold value θ (see 3.2.2) for increasing number of training images, see Fig. 2.a. We fit a low-degree (degree=4) polynomial curve to the data and we finally find the point on the curve which is closest to the origin. T_{tr} is in the range of 300-500 frames for all datasets. The remaining training images ($T_{tr}+1$ to T) are not used in the application phase but for determining the last parameter (K). This is achieved by calculating the prediction error for the images in the remaining training set. We allow $K = 1, 2, ..., 10$, and then choose the K that corresponds to the minimum error. Figure 2.b shows an example of the mean prediction error as a function of K for a representative sequence.



a. Training Phase L-curve



b. Mean prediction error (in mm) vs number of nearest neighbors

Fig. 2. Example of parameter optimization (MRI dataset2). **a.** The number of training images T_{tr} is selected to provide a good trade-off between density and size of the training set using the L-curve method. **b.** K is chosen so that the mean prediction error evaluated over the remaining training images is minimum.

Using the aforementioned parameters, we evaluate our method for each dataset. Results are listed in Table 2. The performance of the method is evaluated comparing the prediction with respect to registration results, see (4). As the registration itself is not without error, we also asses the registration accuracy with respect to the ground truth. On average, the prediction error $\bar{E}$ is less than 1.2 mm

for both MRI (1.17 $\pm$ 0.83 mm) and US images (1.20 $\pm$ 1.40 mm). The average registration error only increases by 0.45 mm for MRI to 2.07 mm and slightly decreases by 0.04 mm for US to 2.07 mm. The percentage of required registrations during the application phase is on average 3.44% (4.44%) for MRI (US) and always below 9%.

The average time needed to obtain a prediction for each application image is 9 ms, using unoptimized MATLAB software and no GPU parallel computing (single PC with Intel®Core™i7-920 at 2.66 GHz processor and 8 GB RAM). Such a performance is faster then the usual acquisition time for both image modalities.

	Registration vs GroundTruth	Motion Magnitude
MRI 1	1.94 $\pm$ 0.81 (3.41)	5.63 $\pm$ 4.43 (14.48)
MRI 2	1.40 $\pm$ 0.77 (2.90)	3.62 $\pm$ 3.03 (8.88)
MRI 3	1.80 $\pm$ 1.23 (3.93)	7.45 $\pm$ 7.76 (21.41)
MRI 4	1.31 $\pm$ 1.45 (2.44)	5.64 $\pm$ 4.71 (15.06)
Mean	**1.62 $\pm$ 1.07 (3.17)**	**5.58 $\pm$ 4.98 (14.96)**
US	2.11 $\pm$ 1.36 (3.94)	6.41 $\pm$ 2.81 (12.33)

Table 1. Mean $\pm$ SD of the registration error (in mm) w.r.t. manually selected landmarks and their motion magnitude. In brackets the 95th percentile of the registration error and the motion magnitude.

	Prediction vs Registration	Prediction vs GroundTruth	Appl.Phase Registrations
MRI 1	0.98 $\pm$ 0.80 (2.26)	1.99 $\pm$ 1.09 (4.26)	2.98 %
MRI 2	0.94 $\pm$ 0.69 (2.26)	1.63 $\pm$ 0.78 (3.06)	1.69 %
MRI 3	0.56 $\pm$ 0.41 (1.31)	1.81 $\pm$ 1.14 (3.70)	8.51 %
MRI 4	2.18 $\pm$ 1.44 (4.89)	2.84 $\pm$ 1.38 (4.91)	0.60 %
Mean	**1.17 $\pm$ 0.83 (2.68)**	**2.07 $\pm$ 1.10 (3.98)**	**3.44 %**
US	1.20 $\pm$ 1.40 (3.08)	2.07 $\pm$ 1.15 (3.95)	4.44 %

Table 2. Results of the proposed method for all datasets. Mean $\pm$ SD of the prediction error (in mm) w.r.t. registration (left column) and manually selected landmarks (middle column). In brackets the 95th percentile of the prediction errors. Percentage of the images that require registration during the application phase (right column).

5. CONCLUSION

We proposed a simple and flexible method to speed-up image registration for image sequences of organs subject to repetitive movement, such as breathing. The method exploits the redundancy of the images in a repetitive motion scenario while still being able to adapt to irregularities. Results show real-time capability even for a non-optimized implementation and accuracy comparable to state-of-the-art methods. Tests on respiratory liver motion show that on average 4% of the images required registration during the application phase, despite acquisitions up to an hour and irregular changes in exhale position (drift) [10]. Furthermore, the number of required training images is relative low ($<$500).

We plan to acquire more image sequences and optimize the size of the training set. We expect the proposed method to be generally beneficial for real-time tracking of mainly repetitive motion (e.g. respiratory, cardiac). In particular, it will substantially improve the real-time capability of intensity-based 3D image registration, which has a considerably higher computational complexity.

6. REFERENCES

[1] P.J. Keall, G.S. Mageras, J.M. Balter, R.S Emery, K.M. Forster, S.B. Jiang, J.M. Kapatoes, D.A. Low, M.J. Murphy, B.R. Murray, C.R. Ramsey, M.B. Van Herk, S.S. Vedam, J.W. Wong, and Yorke E., "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med. Phys.*, vol. 33, pp. 3874, 2006.

[2] H. Shirato, S. Shimizu, K. Kitamura, and R. Onimaru, "Organ motion in image-guided radiotherapy: lessons from real-time tumor-tracking radiotherapy," *International Journal of Clinical Oncology*, vol. 12, pp. 8–16, 2007.

[3] M. von Siebenthal, G. Szekely, A. Lomax, and P. Cattin, "Inter-subject modelling of liver deformation during radiation therapy," in *Proc. MICCAI 2007*. 2007, vol. 4791, pp. 659–666, Springer.

[4] J.R. McClelland, S. Hughes, M. Modat, A. Qureshi, S. Ahmad, D.B. Landau, S. Ourselin, and D.J. Hawkes, "Inter-fraction variations in respiratory motion models," *PMB*, vol. 56, pp. 251, 2011.

[5] H. Shirato, K. Suzuki, G.C. Sharp, K. Fujita, R. Onimaru, M. Fujino, N. Kato, Y. Osaka, R. Kinoshita, H. Taguchi, S. Onodera, and K. Miyasaka, "Speed and amplitude of lung tumor motion precisely detected in four-dimensional setup and in real-time tumor-tracking radiotherapy," *International Journal of Radiation Oncology Biol. Phys.*, vol. 64, no. 4, pp. 1229–1236, 2006.

[6] M. Ries, B. D. de Senneville, S. Roujol, Y. Berber, B. Quesson, and C. Moonen, "Real-time 3D target tracking in MRI guided focused ultrasound ablations in moving tissues," *Magnetic Resonance in Medicine*, vol. 64, no. 6, pp. 1704–1712, 2010.

[7] E.J. Harris, N.R. Miller, J.C. Bamber, J.R.N. Symonds-Tayler, and P.M. Evans, "Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound," *PMB*, vol. 55, pp. 3363–3380, 2010.

[8] P. Foroughi, P. Abolmaesumi, and K. Hashtrudi-Zaad, "Intra-subject elastic registration of 3D ultrasound images," *Med. Image Anal.*, vol. 10, no. 5, pp. 713–725, 2006.

[9] K. Zhang, M. Guenther, and A. Bongers, "Real-time organ tracking in ultrasound imaging using active contours and conditional density propagation," in *MIAR 2010*. 2010, vol. LNCS 6326, pp. 286–294, Springer.

[10] M. von Siebenthal, G. Szekely, U. Gamper, P. Boesiger, A. Lomax, and P. Cattin, "4D MR imaging of respiratory organ motion and its variability," *PMB*, vol. 52, pp. 1547–1564, 2007.

[11] P. Arnold, F. Preiswerk, B. Fasel, R. Salomir, K. Scheffler, and P.C. Cattin, "3D organ motion prediction for MR-guided high intensity focused ultrasound," in *Proc. MICCAI 2011*. 2011, vol. LNCS 6892, pp. 623–630, Springer.

[12] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," in *Med. Imaging, Trans. on*. IEEE, 1999, vol. 18, pp. 712–721.

[13] L.J.P. Van der Maaten, E.O. Postma, and H.J. van den Herik, "Dimensionality reduction: A comparative review," *TiCC TR*, 2009-005.